

BACTERIAL COMMUNITY DYNAMICS AND VARIABILITY IN SHALLOW
AQUIFERS

by

Anna Jesus Zelaya

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2019

©COPYRIGHT

by

Anna Jesus Zelaya

2019

All Rights Reserved

DEDICATION

Para mis queridos padres, Guillermo y Catalina.

ACKNOWLEDGEMENTS

I would first like to thank my advisor Dr. Matthew Fields. I am grateful for the many opportunities to attend scientific conferences, and for the many levels of support he provided throughout the years. I am grateful to all past and present committee members, Drs. Gill Geesey, Anne Camper, Christine Forman and Seth Walk. A very special thank you to Dr. Al Parker for his endless patience through the many hours he spent having conversations with a microbiologist eagerly trying to understand statistics. I would also like to thank Dr. Heidi Smith for her mentorship. Thank you to everyone in the Center for Biofilm Engineering, the Microbiology and Immunology Department, and my past and current lab group members for all of their help throughout the years.

I am grateful to the Molecular Bioscience Program and the U.S. Department of Energy's ENIGMA program for financial and professional support throughout my graduate studies.

I would also like to thank the Department of Education, most notably Drs. Tricia Seifert and Carrie Myers, for their contribution towards helping me discover a passion for education.

I would like to thank my partner Greg, who was a constant source of support. Finally, I would like to thank my loving parents, Guillermo and Catalina, who sacrificed everything to give their children the opportunities that they never had. I could not have done this without them.

TABLE OF CONTENTS

1. INTRODUCTION	1
The Shallow Terrestrial Subsurface Environment	1
Shallow Subsurface Microbial Life	4
Diversity of Shallow Subsurface Microbial Communities	4
Shallow Subsurface Microbial Diversity Across Spatiotemporal Scales	6
Time-Series Analysis in Ecology.....	8
Oak Ridge Reservation	10
Research Goals and Objectives.....	11
2. HIGH SPATIOTEMPORAL VARIABILITY OF BACTERIAL DIVERSITY OVER SHORT TIME SCALES WITH UNIQUE GEOHYDROCHEMISTRY WITHIN A SHALLOW, PRISTINE AQUIFER	13
Contribution of Authors and Co-Authors	13
Manuscript Information Page	15
Abstract	16
Introduction.....	17
Materials and Methods.....	20
Results.....	25
Discussion	43
Conclusion	51
3. POPULATION FILTERING IN SEDIMENT BIOFILMS FROM DYNAMIC, SOURCE PLANKTONIC COMMUNITIES	52
Contributions of Authors and Co-Authors.....	52
Manuscript Information Page	54
Abstract	55
Introduction.....	56
Materials and Methods.....	58
Results.....	63
Discussion	75
4. COMMUNITY STRUCTURE ACROSS PARTICLE SIZE IN-VITRO AND IN-SITU	81
Contributions of Authors and Co-Authors.....	81
Manuscript Information Page	82

TABLE OF CONTENTS CONTINUED

Abstract	83
Introduction	84
Materials and Methods	86
Results	93
Discussion	103
 5. CONTRIBUTION OF TOTAL AND VIABLE DNA POOLS ON DIVERSITY ESTIMATES OF CONTAMINATED AND NON-CONTAMINATED SUBSURFACE SEDIMENTS AND GROUNDWATER	 106
Contributions of Authors and Co-Authors	106
Manuscript Information Page	107
Abstract	108
Introduction	109
Materials and Methods	113
Results	123
Discussion	130
 6. EPILOGUE	 134
 APPENDICES	 142
APPENDIX A: Supplementary Materials for Chapter Two	143
APPENDIX B: Supplementary Materials for Chapter Three	154
APPENDIX C: Supplementary Materials for Chapter Four	158
APPENDIX D: Physiological characterization of environmental isolate <i>Desulfosporosinus</i>	163
APPENDIX E: Impact of hydrologic boundaries on microbial planktonic and biofilm communities in shallow terrestrial subsurface environments	167
 REFERENCES	 208

LIST OF TABLES

Table	Page
2.1. Boxplot statistics of measured groundwater geochemistry per well over weeks	27
2.2. Number of times cosmopolitan family appeared at >1% over 3-month sampling period.....	32
2.3. Number of COG's and potential function.....	39
2.4. Results of VAR model and Granger Causality using the Wald Test for a given lag (p) chosen that eliminated autocorrelation	42
3.1 Number of OTUs sampled in groundwater and surrogate sediment per well	65
3.2 Average sampled groundwater geochemistry for three background field wells over three-month incubation period for surrogate sediments.	65
4.1. Measured Geochemistry over time for field wells sampled in this study	94
A.1. OTUs that contributed to ≥ 30 paired-week contributions in SIMPER analysis.	144
A.2. Estimated residence times based on hydraulic conductivity K of varying orders of magnitude using previous field data.	144
A.3. Most abundant resident OTUs per well and their taxonomic classification	149
B.1. Sequence and OTU counts for groundwater and surrogate sediment	152
B.2. List of sampled Phyla in groundwater and surrogate sediment and proportion of samples in which they were observed	152

LIST OF TABLES CONTINUED

Table	Page
B.3. OTUs most strongly associated with surrogate sediment based on CCA	154
C.1. Number of beads per draw for 30-50um particles.....	157
E.1. Gram reaction and cell size comparisons of <i>Desulfosporosinus</i>	163
E.2. Types of media used for D1 isolate growth.....	163

LIST OF FIGURES

Figure	Page
1.1 Diagram of shallow terrestrial subsurface and aquifer with unsaturated, saturated, and transition zones.....	3
2.1 Tukey-style boxplots of within-well geochemistry	26
2.2 A) Tukey-style boxplots of measured within-well ground-water diversity.....	28
B) Plots of weekly means for within-well groundwater diversity.....	29
2.3 A) 20 cosmopolitan taxa observed in groundwater samples.	30
B) Transiently abundant taxa observed in >1% abundance.....	31
2.4 Bubble plots of relative abundances for cosmopolitan groups.....	35
2.5 A) Unconstrained ordination (DCA) of OTUs per sample	
B) Samples ordinated based on Functional Gene Array	36
2.6 Functional Gene Array (Geochip) results	38
2.7 Multivariate ordination of community structure and geo-chemistry	41
2.8 Association networks based on VAR models.....	41
2.9 Relative influence of community assembly processes.....	43
3.1 Venn Diagrams of total OTUs in sampled groundwater and surrogate sediment	64
3.2 Diversity profile of sampled groundwater and surrogate sediment	67
3.3 A) Stacked relative abundance bar graphs for 11 shared surrogate sediment families and	
B) their corresponding groundwater abundances	69

LIST OF FIGURES CONTINUED

Figure	Page
3.4 A) Most abundant OTUs in surrogate sediments and corresponding abundances in filtered groundwater, and B) Most abundant OTUs in groundwater and their corresponding abundances in surrogate sediment.....	71
3.5 Representation of pairwise Bray-Curtis distances between groundwater and surrogate sediment communities using NMDS	72
3.6 Community networks of co-occurring OTUs based on correlation analysis of within-well groundwater and sediment populations with SPARCC	74
3.7 Relative influence of community assembly processes on spatiotemporal turnover of within-well assemblages	75
4.1 Boxplots of (A,B) Richness and (C,D) Effective Species Number per bead size.....	96
4.2 Representation of pairwise Bray-Curtis distances between samples using two-dimensional NMDS.....	96
4.3 Richness and Diversity estimates based on Hill Numbers.....	97
4.4 Diversity estimates based on Hill Numbers per well.....	98
4.5 Stacked relative abundance bar graphs	99
4.6 Percent fraction of rare, intermediate, and abundant OTUs residing on particles of different sizes per environment type	100
4.7 Heatmap of top 1% OTUs.....	102
5.1 Cross sectional view of sampled sediment cores (EB271 and EB106) and groundwater wells (GW271 and FW106) from EOR-FRC.....	114
5.2 Viability PCR workflow using PMA	123

LIST OF FIGURES CONTINUED

Figure	Page
5.3 Results of initial PMA tests with <i>E. coli</i> cultures	124
5.4 A) qPCR results of spiked <i>E. coli</i> DNA experiments.....	125
B) Boxplots representing the range of amplified 16S rRNA gene copies	125
5.5 Diversity measures of PMA and non-PMA treated ground- water.....	127
5.6 Stacked relative abundance bar graph of average relative abundances of PMA treated and non-treated samples per sample type	128
5.7 Bar plots of relative abundances of OTUs in PMA treated an untreated cored sediment per depth	129
A.1 Geographic location of non-contaminated background site of the EOR-FRC and chosen field wells	140
A.2. (a-d) Tukey-style boxplots of within-well geochemistry.....	140
(e-g) Scatter plots of measured organic acids over time.....	141
A.3. Plots showing weekly means of within-well groundwater geochemistry	141
A.4. Relative abundances by a) Phylum and b) lowest classifiable taxa.....	142
A.5. Minimum, average, and maximum dissimilarities of OTU diversity across wells	143
A.6. Relative abundances per sample for OTUs that contributed $\geq 1\%$ in at least one paired-week Bray-Curtis dissimilarity based on SIMPER	143
A.7. Relative abundance bar charts for resident and transient OTUs observed in FW301	145

LIST OF FIGURES CONTINUED

Figure	Page
A.8. Relative abundance bar charts for resident and transient OTUs observed in FW303	146
A.9. Relative abundance bar charts for resident and transient OTUs observed in FW305	147
A.10. Detrended correspondence analysis based on observed resident OTUs for each well	148
A.11. Average pair-wise Bray-Curtis dissimilarities of resident OTUs over time per well.....	150
B.1. Surrogate sediment samplers and deployment schematic	151
B.2. Relative abundance plots of Phyla observed per sample	152
B.3. Venn-Diagram of shared OTUs groundwater and surrogate sediment	153
B.4. Canonical Correspondence Analysis (CCA) of surrogate sediment communities and associated geochemistry on the day they were harvested.....	153
C.1. Boxplots of richness per bead size and environment type	155
C.2. Boxplots of Effective Species Numbers per bead size and Environment type.....	156
C.3. Laboratory bead reactor setup	157
C.4. Total cell counts and organic matter content for differently sized beads incubated in laboratory reactors	158
C.5. EEAMs spectral characterization	159
D.1. Gram- and fluorescent stains of D1 isolate	161
D.2. Phylogenetic relatedness of unknown D1 isolate	162
D.3. Growth curves of D1 isolate	162

LIST OF FIGURES CONTINUED

Figure	Page
E.1. Conceptual illustration of representative shallow subsurface environment that includes the vadose, capillary fringe, and saturated zones.....	174
E.2. Conceptual model of subsurface flow and mixing zones and potential effects on biofilm life-cycle dynamics	186

ABSTRACT

The shallow, terrestrial subsurface plays an important role in sustaining life above ground. Globally, subsurface environments are becoming increasingly threatened by anthropogenic sources of contamination and disturbance. The Oak Ridge Reservation (ORR) played an important role in the development of nuclear weapons during World War II. In the process, the underlying subsurface became contaminated with radioactive and hazardous wastes. Strategies to remediate polluted subsurface environments such as Oak Ridge have included natural attenuation by resident microbes such as bacteria. Such studies would be aided by a thorough understanding of the natural variability of microbial diversity over space and time in uncontaminated environments. We examined microbial community diversity and variability in both contaminated and non-contaminated sites of the ORR via ss-rRNA paired-end sequencing. An initial *in-situ* spatiotemporal survey of non-contaminated groundwater was conducted in order to understand how diversity changes over time in an undisturbed aquifer. Moreover, different models were used to estimate possible causal relationships between geochemical parameters and population distribution. Additionally, surrogate sediment samplers were filled with native sediment to assess the diversity of the attached bacterial fraction. Communities observed in the attached fraction were a subset of groundwater communities, although the dominant fractions of each were distinct. After initial assessment, *in-vitro* groundwater was used as both a nutrient source and microbial inoculum for bioreactors filled with glass beads of various sizes in order to understand particle size effects on community dynamics. Potential viability of resident microbes in both contaminated and non-contaminated groundwater and cored sediments was assessed using PMA-Seq and other complementary methods. Collectively, results show that 1) microbial communities in groundwater are highly dynamic over short time-scales, 2) corresponding changes in geochemistry are mostly weakly related to changes in community structure (except perhaps after a disturbance or stress period), 3) community assembly may be affected by pore space volume, 4) PMA-viable populations differ between solid and aqueous fractions, 5) most subsurface cells are intact, and 6) traditional estimates of largely abundant populations may be influenced by the presence of DNA from non-viable members, resulting in less abundant populations being unmeasured or underestimated. The results of this study have implications for sampling and appropriate estimations of microbial populations *in situ* as well as the inherent variability in an uncontaminated shallow aquifer.

CHAPTER ONE

INTRODUCTION

Terrestrial, shallow subsurface environments are globally important but poorly understood ecosystems. Subsurface environments provide valuable resources that sustain life above ground, such as clean and potable freshwater (Gleick et al., 2018), geothermal energy (Lee, 2013) and nutrient rich soils for agriculture (Manale et al., 2018). Increased demands for these resources, as well as historically poor management of many subsurface environments, have led to widespread contamination of groundwater and subsurface soils and sediments. Subsurface research by geologists, hydrologists, engineers, and ecologists has been largely motivated by the necessity to better understand and manage this valuable environment (Griebler and Leuders, 2009; van Ree and van Beukering, 2016). Additionally, the potential for natural remediation (*i.e.*, attenuation) of disturbed subsurface environments by native microorganisms has raised interest in subsurface microbial ecology (Holliger et al., 1997; Chakraborty et al., 2012).

The Shallow Terrestrial Subsurface Environment

Located beneath vertically weathered surface soil layers, the shallow subsurface is comprised of sediments, rocks, gas, pore water, and groundwater (Atekwana et al. 2006, Fig. 1). Although estimates vary, the shallow subsurface environment can extend from directly beneath the organic-matter rich soil layers (A and B horizons) to tens of meters deep until reaching bedrock (Atekwana et al. 2006; Pepper & Brusseau 2006; Chu et al.

2016). The degree of hydrological connectivity to the surface (Toth 1963) is often used to differentiate shallow versus deep subsurface environments. Shallow subsurface environments have also been described as continental subsurface environments that are composed of the earth's crust, mineral ores, and aquifer environments (Jones et al., 2018).

Shallow subsurface aquifers are typically described as having three distinct zones: vadose, capillary fringe, and saturated (Appendix F, Fig F.1). Typically, the presence of water and vertical distance to the water table are used to distinguish between zones. Water enters the subsurface via the porous and transiently saturated vadose zone, percolating downwards towards the water table. The zone at or below the water table is termed the saturated zone and consists of porous parent material with voids that are permanently filled with water. The intermediary zone between the vadose and saturated zones is the capillary fringe and can display high levels of physiochemical fluctuations resulting from water table changes (Griebler & Lueders 2009). This fluctuating interface has been shown to be a 'hotspot' of subsurface activity, especially with respect to biogeochemical cycling (Silliman et al. 2002; Berkowitz et al. 2004).

The shallow terrestrial subsurface is characterized by a heterogeneous assemblage of materials forming a porous medium where solid, liquid, and gaseous phases are present. Solid phases of the subsurface are composed of weathered products from parent material at various stages of development. The mineral content and physicochemical properties of underlying parent rock material have a large influence on the texture, porosity, and permeability of overlying weathered sediments and surface soils (Berkowitz et al., 2004).

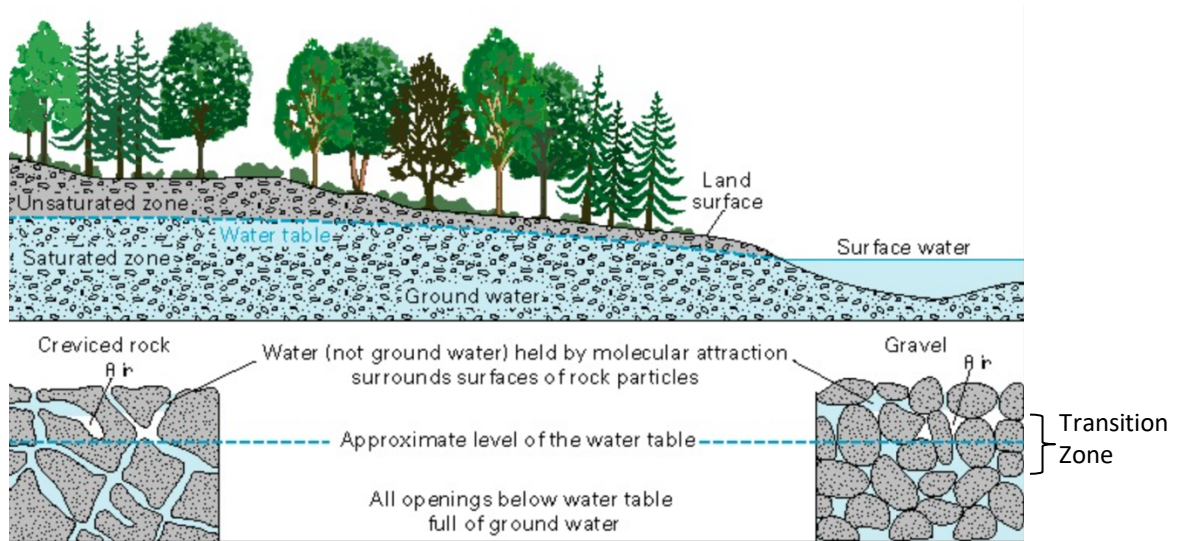


Figure 1.1 Diagram of shallow terrestrial subsurface and aquifer with unsaturated, saturated, and transition zones. The unsaturated zone above the water table may contain water within pore spaces, but it is not fully or completely saturated with water. The saturated zone lies below the water table and is distinguished by full saturation of solids and pore spaces. The zone directly at the water table is transiently saturated often due to capillarity as a result of fluctuating water levels of the water table and is generally referred to as the capillary fringe. The two drawings at the bottom of the diagram show a close-up of how water is stored in between underground rock particles. Original figure modified from Waller, 1988.

The unique characteristics of water (*e.g.*, polarity, cohesion, adhesion, and surface tension), allow it to dissolve solutes, attach to subsurface solids, fill pore-spaces, and move through the subsurface (Brady and Weil, 2014). Subsurface water geochemistry (*e.g.*, pH, salinity, trace elements, metals) is largely dependent on the surrounding geology, residence time, and subsurface usage history (Brady and Weil, 2014). Additionally, the surrounding geology influences water movement and transport characteristics. For example, the surface area of different mineral types directly effects retention rates, solid-water adhesion, and water-air ratio within pore-spaces (Berkowitz, 2014). Hydrological flow rates can be difficult to measure directly (National Research Council, 2000). While natural fluctuations

in groundwater movement and composition do occur (Nelson, 2002), short- and long-term fluctuation rates, the variability across different locations within the same aquifer system, and the ecological effects of those fluctuations, are less clear. In terms of varying hydrology and water infiltration, altered flow conditions and water activity can limit cellular motility/dispersal as well as nutrient availability that can result in decreased microbial activity and altered population distribution(s) (Or *et al.* 2007).

Shallow Subsurface Microbial Life

Shallow subsurface environments harbor a rich and diverse array of microbial life. Traditional culturing and microscopic techniques provided the first clues towards microbial richness within the subsurface (Balkwill, 1997) and continue to provide important insights into microbial physiology and phenotypes. Technological advances in molecular methods (*e.g.*, molecular surveys using the 16S rRNA gene, metagenomics, and targeted functional genes) have been instrumental in detecting far more taxonomic and functional diversity than was previously believed to exist (Brown *et al.*, 2015; Lynch and Neufeld, 2015; Lennon and Locey, 2016). Despite these advances, many aspects of how microbial communities form, function, and vary over space and time is still not well understood, particularly for difficult to sample environments such as the subsurface.

Diversity of Shallow Subsurface Microbial Communities

While specific taxonomic lineages can be prevalent in several types of underground ecosystems (Griebler and Lueders 2009; Akob and Küsel 2011; Hubalek *et al.* 2016), thus far no true ‘endemic’ shallow subsurface populations have been identified

(Griebler and Lueders 2009). It has been recently shown that subsurface groundwaters and sediments contain a variety of previously undescribed bacteria and archaea (Castelle et al. 2013; Brown et al. 2015; Anantharaman et al. 2016; Lazar et al. 2017). Interestingly, it has also been shown that molecular survey techniques that depend on abundances of nucleic acid markers (*e.g.*, SSU rRNA) may overrepresent certain members within subsurface soils due to the co-extraction of DNA from extracellular pools or from cells that are dead, dying, or with compromised membrane integrity (Carini et al., 2016). However, these observations may not be universal but rather dependent on the type of environment under study (Ramirez et al., 2018). Community assembly processes that affect global microbial distributions have been hypothesized based on surface habitat observations (*i.e.*, biotic interactions, dispersal limitation, and environmental filtering) (Martiny et al. 2006; O'Malley 2007; Griebler & Lueders 2009; Shoemaker et al. 2017). However, it remains unclear whether these mechanisms hold true for the distribution of microorganisms within the oligotrophic subsurface (Musslewhite et al. 2003; Chu et al. 2016).

Historically, microbial surveys have typically been performed on more easily accessible vadose soils. However, the hydrogeologic, geochemical, and geophysical constraints of the vadose zone are not entirely representative of lower depths. Microorganisms residing in the layers below the vadose experience a different set of constraints. For example, unlike surface soils, the lack of light precludes photosynthesis as a viable mode of life (Freckman et al., 1997). However, organic carbon does infiltrate the lower depths through infiltration over the watershed scale. Dissolved oxygen also typically decreases with depth and can vary with infiltration and microbial activity (Nimmo, 2004).

As such, microorganisms with abilities to utilize alternative and less thermodynamically favorable electron acceptors are observed in greater proportions across vertical depth (Flynn *et al.*, 2013). Currently, it remains unclear if this trend is merely a consequence of the combination of limited/recalcitrant resources (C) and energy restriction (anoxic) or other more specific selection mechanisms that may differ from surface environments (Musslewhite *et al.* 2003; Lin *et al.* 2012b; Chu *et al.* 2016). Recent studies suggest that the large fraction of lowly abundant or ‘rare’ organisms observed in subsurface environments may play important ecological roles. For example, they may contribute to biogeochemical reactions (Pester *et al.* 2010) while also serving as a ‘microbial bank’ that can ‘seed’ environments when conditions change (Lynch and Neufeld 2015). Biofilms could play a major role across the shallow subsurface zones in which changing conditions (*e.g.* pH, conductivity and flow) could drive dispersal and/or invasion.

Shallow Subsurface Diversity Across Spatiotemporal Scales

Across ecosystem spatial scales, there is limited information regarding the exact relationship between environmental and biogeochemical constraints on microorganisms within subsurface environments. Traditionally, bulk measurements have been used to assess microbial richness, community structure, and activity (Smith *et al.*, 2018). However, the heterogeneous and dynamic habitats of the subsurface (*e.g.*, dynamic fluxes of water, nutrients, temperature, pH, and osmolarity) may create a variety of differing microhabitats at spatial scales relevant to microorganisms (*e.g.*, micrometers, millimeters).

It is estimated that half of the total volume of subsurface environments is comprised of solid materials, and the other half is made up of liquid and gas phases (Alexander, 1977). Porosity, pore-size, and permeability help to describe the physical and geochemical spaces that create macro- and micro-transition zones between the solids, liquid, and gas phases of individual particles within the subsurface. It is hypothesized that such transition zones are important ecotones or ‘hotspots’ of microbial diversity and activity (Zhang et al. 1998; McClain et al. 2003; Goldscheider et al., 2006; Bougon et al., 2012; Campbell et al., 2012; Jones et al., 2014) and deserve more careful attention. Porosity, pore-size, and permeability can all have direct influence over the movement and transport of liquids, gases, and nutrients (Hay et al., 2011; Rabbi et al., 2015), the presence and type of microorganisms that can reside in individual pore-spaces (Frederickson et al., 1997; Ruamps et al., 2011), reaction rates (Liu et al., 2015), microbial growth rates (Coyte et al., 2016), and redox-potential (Briggs et al., 2015). While sampling and analyzing environments across multiple scales (*e.g.* km, m, μm , nm, etc) can be challenging, such studies have provided evidence of spatial (vertical and horizontal) taxonomic variation within groundwater (Lin et al. 2011; Lin et al. 2012a; Herrmann et al. 2015) and sediments (Lin et al. 2012b). Results typically show a decline in microbial richness and diversity over vertical depth. The extent that microbial communities vary in relation to depth, even with application of newer sequencing technologies, is still poorly resolved for the variety of geological strata that represent the shallow subsurface. Therefore, increased spatial resolution is needed to better understand the implications of micro-scale heterogeneity on microbial population distributions over space and time.

Taxonomic diversity in relation to temporal fluctuations in physicochemical and geochemical properties of the shallow subsurface is largely understudied. Physicochemical and geochemical relationships over time may drive phylogenetic and functional microbial diversity changes that are often associated with both short-term (*e.g.*, diel cycles; extreme weather) and long-term (*e.g.*, seasonal) alterations. Spatiotemporal fluctuations in conjunction with porosity and permeability changes can lead to varying degrees of hydrogeochemical mixing. The role of mixing in aquatic systems has been implicated in the creation of patchy distributions of both nutrients and biomass (Ebrahimi & Or 2016; 2018). Depending on the aquifer system, highly mixed waters can be located in shallow areas, where there are higher rates of infiltration from rain or surface waters. Alternatively, at the water table, seasonal fluctuations can result in mixing groundwater with sediments of the variably saturated zone (Appendix E , Figures E.1 and E.2). In the shallow subsurface, mixing is thought to cause instability due to faster and shorter local flow paths; whereas, more stable and predictable diversity may result from slower regional flow in deeper aquifer systems (Ben Maamar et al. 2015).

Time-Series Analysis in Ecology

How communities change over space and time, and how past changes may help to predict future ones, have been central questions in ecology and environmental studies. Such ecological studies often involve the sequential collection of samples or recording of observations over space and/or time. Sequentially collected observations can be challenging to analyze using traditional statistical methods that assume independent, random sampling and statistical independence of errors (Bence, et al., 1995). Data collected

over time is often serially-dependent, or autocorrelated. That is, the value of an observation at a given point in space or time is not truly random or independent but instead depends on the preceding values (Borcard et al., 2011). When the strength of that dependence is high, attempting to disentangle the possible influences of environmental factors on the value observed becomes challenging (Borcard et al., 2011). In order to avoid spurious statistical inferences, it is recommended that autocorrelation be identified and accounted for in the statistical models employed (Bence et al., 1995; Hefley et al., 2016).

Time series analyses that account for autocorrelation can be used to detect stability, trends, seasonality and periodicities within an environmental system. Time series can also be paired with cross-predictive techniques (*e.g.*, Granger Causality), which can be used to assess how well the values of one time-series can be used to predict the values of another (Faust et al., 2015). These methods can also be used in the forecasting and prediction of future changes in ecosystem dynamics. Historically, long-term time series (*e.g.*, hundreds of observations over several years) have been preferred for predictive forecasting (Borcard et al., 2012). However, short-term time series (*e.g.*, 10-50 observations over days, weeks, or months) can capture short-term dynamics that might be missed with wider sampling ranges of long-term studies (Fuhrman et al., 2015). Recently, the importance of using both short-term and long-term temporal studies in order to meet the current pace of environmental change and timescales at which environmental policy/management decisions are made (*e.g.*, days, weeks) has been emphasized (Dietze et al., 2018; Stegen, 2018).

Oak Ridge Reservation

The Oak Ridge Reservation (ORR), located in Oak Ridge, Tennessee, is a government operation that consists of three major facilities: The Y-12 National Security Complex, Oak Ridge National Laboratory, and East Tennessee Technology Park. The ORR was first established during the Manhattan Project of the 1940's. It was originally built to process radioactive and non-radioactive materials for nuclear weapons development. After the end of the war, ORR continued to run a variety of research and production activities. Over the course of 40 years of such operations, significant amounts of radioactive, hazardous, and mixed wastes were disposed at several sites throughout the reservation, leading to widespread contamination in the area. In 1989, the Environmental Protection Agency designated the Oak Ridge Reservation as a Superfund site (epa.gov) mandating a long-term clean-up effort. Current research involves a multifaceted approach towards understanding the fate and transport of contaminant materials, as well as the effects on the surrounding biota. Understanding the impacts of ecological disturbances, such as contamination, on subsurface microbial communities and the potential to naturally remediate contaminated areas has also been an intense area of research at the ORR. The availability of a non-contaminated background site with similar underlying geology and mineralogy located in the West Bear Creek Valley, approximately 2 km from any contaminated areas of the ORR, is useful for comparisons of the effect of contamination on native microbial communities. Dozens of monitoring wells have been installed in the non-contaminated background area to access shallow groundwater (20ft - 200ft below the surface) for hydrologic and biological field site studies.

Research Goals and Objectives

1. *Determination of the variability over space and time of non-contaminated groundwater microbial diversity and geochemistry.*
2. *Comparison of source diversity (groundwater) and local diversity (attached) in a non-contaminated shallow aquifer.*
3. *Investigate the role of porosity and particle size on microbial diversity under both in vitro and in situ conditions.*
4. *Assess the potential contribution of extracellular DNA pools on abundance and diversity estimates of bacterial populations.*

Objective 1 is described in Chapter Two where we sampled three shallow field wells (FW301, FW303, FW305) tapping into non-contaminated groundwater in the Oak Ridge Field Research Center (ORR) (Oak Ridge, TN) approximately three times a week over a period of three months to measure changes in groundwater geochemistry and microbial diversity over time.

Objective 2 is described in Chapter Three where we incubated surrogate sediment samples for three months in the same three wells studied in Chapter Two (FW301, FW303, FW305) within the background site of the ORR to determine the inter- and intra-well variation of sediment associated communities. Additionally, local sediment biofilm communities were compared to those of the groundwater (source diversity).

Objective 3 is described in Chapter Four where we incubated glass beads of varying sizes both *in-situ* and in laboratory bioreactors fed with unfiltered groundwater from the background area of the ORR. Community structure that formed on individual beads was compared via 16S rRNA sequencing and analysis.

Objective 4 is described in Chapter Five where we attempted to delineate the viable and active populations with unfiltered groundwater and freshly cored sediments from both contaminated and non-contaminated environments using culture-independent methods.

CHAPTER TWO

HIGH SPATIOTEMPORAL VARIABILITY OF BACTERIAL DIVERSITY
OVER SHORT TIME SCALES WITH UNIQUE GEOHYDROCHEMISTRY
WITHIN A SHALLOW, PRISTINE AQUIFER

Contribution of Authors and Co-Authors

Manuscript in Chapter Two

Author: Anna J. Zelaya

Contributions: Performed literature review, synthesized ideas, performed statistical and bioinformatical analyses, wrote and edited the manuscript

Co-Author: Albert E. Parker

Contributions: Statistical analyses, reviewed and edited the manuscript

Co-Author: Kathryn L. Bailey

Contributions: Performed field samplings, geochemical analysis and data collection, reviewed and edited the manuscript

Co-Author: Ping Zhang

Contributions: Performed DNA extraction and paired-end sequencing, wrote methods, reviewed and edited the manuscript

Co-Author: Joy Van Nostrand

Contributions: Performed DNA extractions, paired-end sequencing, and functional gene array, wrote methods, reviewed and edited the manuscript

Co-Author: Ning Daliang

Contributions: Performed β MNTD and β NTI analysis, wrote methods, analysis reviewed and edited the manuscript

Co-Author: Dwayne A. Elias

Contributions: Supervisor and Principal Investigator to Kathryn L. Bailey, reviewed and edited manuscript

Co-Author: Jizhong Zhou

Contributions: Supervisor and Principal Investigator to Daliang Zhou, reviewed and edited manuscript

Co-Author: Terry Hazen

Contributions: Synthesized ideas, reviewed and edited the manuscript

Co-Author: Adam P. Arkin

Contributions: Synthesized ideas, reviewed and edited the manuscript

Co-Author: Matthew W. Fields

Contributions: Principal Investigator, Synthesized ideas, wrote and edited the manuscript

Manuscript Information Page

Anna J. Zelaya, Albert E. Parker, Kathryn L. Bailey, Ping Zhang, Joy Van Nostrand,
Ning Daliang, Dwayne A. Elias, Jizhong Zhou, Terry C. Hazen, Matthew W. Fields

Status of Manuscript:

____ Prepared for submission to a peer-reviewed journal

__X__ Officially submitted to a peer-reviewed journal

____ Accepted by a peer-reviewed journal

____ Published in a peer-reviewed journal

ABSTRACT

It is estimated that over 98% of the earth's consumable and available freshwater is beneath the earth's surface as groundwater (Ganoulis and Fried, 2018), and microbial communities are responsible for cycling a majority of the organic matter and minerals within subsurface environments. Understanding microbial community structure and function within the subsurface is critical to assessing overall quality and maintenance of groundwater; however, the factors that determine microbial community assembly, structure, and function in groundwater systems and their impact on water quality remains poorly understood. In this study, three shallow wells (FW301, FW303, FW305) in a non-contaminated shallow aquifer in the ENIGMA-Oak Ridge Field Research Center (Oak Ridge, TN) were sampled approximately three times a week over a period of three months to measure changes in groundwater geochemistry and microbial diversity. The three wells ranged in average dissolved oxygen from 0.83 (sd= ± 0.4) to 1.7 (sd= ± 0.5) mg/L, average pH from 6.4 (sd = ± 1.1) to 7.3 (sd= ± 0.6), average conductivity from 173 (sd= ± 68.7) to 303 (sd= ± 26.7) $\mu\text{S}/\text{cm}$, average sulfate from 46 (sd= ± 9.7) to 66 (sd= ± 11.3) μM , and average nitrate from 1.6 (sd= ± 1.3) to 3.8 (sd= ± 1.1) μM . The wells displayed geochemical variability over time unique to each well, with FW303 being overall the most stable well and FW301 being the most dynamic based upon dissolved oxygen, conductivity, and nitrate. Community analysis via ss-rRNA paired-end sequencing and distribution-based clustering revealed higher OTU richness, diversity, and variability in groundwater communities of FW301 than the other two wells for diversity binned over all time points. Microbial community composition of a given well was on average >50% dissimilar to any other well at a given

time (days), yet, functional gene diversity as measured with GeoChip remained relatively constant. Similarities in community structure across wells were observed with respect to the presence of 20 taxa (family level) that were observed on all dates sampled across all three wells. Despite their consistently observed nature, there was still variability as they were observed at fluctuating relative abundances over the tested time period. Similarity percentage (SIMPER) analysis revealed that variability in FW301 was largely attributed to low abundance, highly-transient populations, while variability in the most geochemically stable well (FW303) was due to fluctuations in more highly abundant and frequently present taxa. Additionally, the youngest well FW305 showed a dramatic shift in community composition towards the end of the sampling period that was not observed in the other wells, suggesting possible succession events over time. Time-series analysis using vector auto-regressive models and Granger-causality showed unique relationships between richness and geochemistry over time in each well. These results indicate a highly dynamic system over short time scales, with day-to-day population shifts in local community structure influenced by available source community diversity and local groundwater geochemistry.

Introduction

Groundwater is an essential resource used by humans for drinking, agriculture, and industrial processes such as mining and energy production (Danielopol et al., 2003; Reilly et al., 2008). Currently, global groundwater resources are being threatened by overuse, depletion, and contamination from a variety of sources including agricultural runoff (*i.e.*,

nitrates and pesticides), industrial processes, pharmaceuticals, personal care products, landfill leaching, and improperly managed wastewater treatment systems (Danielopol et al., 2003; Gerth et al., 2004; Sui et al., 2015). Additionally, subsurface ecosystems are increasingly being disturbed via engineering efforts (*e.g.*, managed aquifer recharge systems (Lee and Lee, 2017), geothermal/cooling (van der Gun et al., 2016)), often without a thorough investigation of the microbial ecological networks that may be perturbed. Understanding the role of microbial communities in undisturbed, pristine environments can help assessment of anthropological impacts on subsurface ecosystems and thus inform proper management of water resources.

While subsurface hydrology of groundwater aquifers has been well studied by hydrologists and hydrochemists (Chapelle, 2000; Griebler and Avramov, 2015), the microbiology and ecology of groundwater systems have been relatively less studied (Griebler and Leuders, 2009). Interest in groundwater microorganisms and associated ecological function has increased largely because of the emerging need to assess groundwater quality and fitness for human use, for example determining possible vectors for disease or microbial potential to naturally attenuate contaminated water systems (Anderson & Lovley, 1997; Chapelle, 2000; Langwaldt and Puhakka, 2000; Fields et al., 2005). Other studies have uncovered the significant roles that microorganisms play in groundwater processes, such as water geochemistry (Chapelle, 2000), rates of mineral weathering in aquifers (Akob and Küsel, 2011), and the fate and transport of metals and organic compounds (Anderson and Lovley, 1997; Fields et al., 2006; Hwang et al., 2009). While much of the attention given to shallow aquifer systems has been on the effects of

contamination and other disturbances (Fields et al., 2005; Zinger et al., 2012; references therein), knowledge of the activity and composition of microorganisms in pristine groundwater environments have been far less studied (Sogin et al., 2006; Griebler and Leuders, 2009, Caporaso et al., 2010; Zinger et al. 2012), and understanding of the spatial and temporal dynamics in groundwater environments is still in its early stages (Smith et al., 2018).

The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA)-Oak Ridge Field Research Center (EOR-FRC), established by the U.S. Department of Energy (DOE), consists of several contaminated areas as well as a non-contaminated background site. Multiple field wells in the background area access a shallow, oligotrophic aquifer. To the authors' knowledge, there has not been an exploration of the spatial and temporal stability of the microbial diversity in the EOR-FRC pristine aquifer. A clearer understanding of the natural variability in community structure in pristine aquifers is necessary to fully assess the impact of disturbances on aquifer environments. The described study analyzed the spatial and temporal variability of microbial communities in a pristine aquifer at the EOR-FRC and sought to answer the following questions: (1) How similar is the geochemistry of three wells in an undisturbed aquifer system located in close proximity to one another over time? (2) How stable are the microbial communities within and across these wells over time? (3) Does potential functional capabilities of the microbial communities remain stable over time? (4) What is the relationship between the groundwater geochemistry and microbial community composition over space and short-time periods.

Materials and Methods

Site Location

The EOR-FRC background site covers an area of approximately 1.63 km² and is located in the Tennessee West Bear Creek Valley, about 2 km away from any contaminated areas (Appendix A, Supplementary Fig. A.1). No known contaminants have been disposed at the background area of the EOR-FRC throughout the history of DOE operations. The subsurface is comprised of underlying bedrock, interbedded shale and limestones, which weather into unconsolidated and lowly-permeable clay-rich saprolite (Watson et al., 2004). The saprolitic layer in turn is overlain by approximately 0.5 to 3 m of organic and highly permeable clay-rich soil (Solomon et al., 1992). The water table underneath the chosen field wells is considered shallow, lying approximately 6.6 m below the subsurface. Vertical percolation of groundwater occurs in the upper porous layers, however hydrological studies of the area show that preferential flow paths are created along fractures in the underlying rock. Tracer studies reveal that the flow paths are poorly connected in three dimensions (Solomon et al., 1992). Therefore, while the three field wells chosen for this study tap into the same pristine aquifer, it is concluded that the three wells lie along separate, unconnected flow paths within the same watershed.

Groundwater Sampling

The three field wells chosen were FW-301, FW-303, and FW-305 and are considered shallow wells (<17 m depth) (Appendix A, Supplementary Fig. A.1). Filtered groundwater samples were collected approximately three times a week over approximately a three-month period (July- October) after two well volumes were purged. Groundwater

was pumped from the wells and successively filtered using a 10µm filter (to catch large particulates) followed by a 0.2 µm PES filter to collect biomass for nucleic acid extraction. Filters were stored at -80°C until time of nucleic acid extraction. In an attempt to analyze equidistant samples, 61 of the 84 samples were chosen for paired-end barcoded sequencing (~20 samples per well). Additionally, a total of 12 environmental variables were measured (temperature, pH, oxidation/reduction potential, dissolved oxygen, conductivity, lactate, acetate, propionate, fluoride, chloride, nitrate, and sulfate) for every sampling time point. Precipitation data (rainfall) were also monitored and recorded daily. Lactate was measured below detection for a majority of samples (supplementary material); therefore, it was dropped from subsequent statistical analyses.

Nucleic Acid Extraction, Sequencing, and Downstream Analysis

Nucleic acids were obtained using the modified miller method of nucleic acid extraction (Hazen et al., 2010). The SSU rDNA gene sequences were amplified using a primer pair targeting the V4 region (Forward primer, 515F, 5'-GTGCCAGCMGCCGCGGTAA-3'; reverse primer, 806R, 5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso et al., 2012). PCR amplification was performed in triplicate and prepared for sequencing according to the MiSeq™ Reagent Kit Preparation Guide using a 500-cycle v2 MiSeq reagent cartridge (Illumina, San Diego, CA, USA).

Raw sequence data was processed using an in-house pipeline (E. Alm, <https://github.com/almlab/SmileTrain/wiki>). Reads were merged, removed of primers, quality checked, de-replicated, and checked for chimeras. USEARCH was used to merge

reads, filter low quality reads, and remove chimeras (Edgar, 2013). Operational Taxonomic Units (OTUs) were clustered using Distribution Based Clustering (Preheim et al., 2013). Taxonomic annotation of individual OTUs were obtained with the RDP 16S Classifier (Wang et al., 2007). Only classifications with an estimated confidence of >80% were used in downstream analyses. The OTUs below this cutoff were assigned as “unidentified Bacteria.” Additionally, sequences classified as “Chloroplast” were removed from dataset. The resulting classified OTU data table was further filtered to remove rare OTUs. Rare OTUs were defined as OTUs that had a relative abundance of $<10^{-6}$ across all groundwater samples (Haegeman et al., 2013). The representative sequences were used to build phylogenetic tree by FastTree2 (Price et al., 2009, Price et al., 2010) after aligned by Clustal Omega (version 1.2.2) (Sievers et al., 2011).

Data Analysis

Initial diagnostics (normality, residual plots) were performed using R statistical package (R Core Team, 2017). Richness and Diversity using Hill Numbers as well as Sorensen and Bray-Curtis pair-wise dissimilarities were determined in R with package VEGAN (Oksanen et al., 2013). Correlation analysis and linear regression models were used to identify relationships among the geochemical and microbial data using lme4 package (Bates, 2010). Analysis of Variances (ANOVA) was performed to determine mean differences across weeks within each well and across wells. Constrained (Canonical Correspondence Analysis, or CCA) and unconstrained (Detrended Correspondence Analysis, or DCA) ordinations were carried out in R statistical package with VEGAN. Similarity percentage (SIMPER) analysis using PAST statistical package (Hammer et al.,

2001) was used to determine specific OTUs contributing to weekly pair-wise dissimilarities.

GeoChip Hybridization and Analysis

GeoChip hybridization was performed as described with aliquots (5-10 ng for measurable DNA or 2-5 μ l if no measurable DNA) of DNA amplified using the Templiphi kit (GE Healthcare) with the following modifications. Spermidine (0.1 mM) and single stranded binding protein (267 ng ml⁻¹) were added to improve the amplification efficiency and samples were amplified for 6 h (Wu et al., 2006). Amplified DNA (~2 μ g) was labeled with mixed cyanine 3 using random primers (Life Technologies, random hexamers, 3 μ g/ μ L) and Klenow (imer; San Diego, CA; 40 U mL⁻¹) at 37°C for 6 h and then cleaned using a QIAquick purification kit (Qiagen) per the manufacturer's instructions and dried. Labelled DNA was rehydrated with hybridization buffer and hybridized for 20-22 hr at 67°C and then imaged (NimbleGen MS 200 microarray scanner) and data was extracted using the Agilent Feature Extraction program. Data normalization and quality filtering were performed with multiple steps (Liang et al., 2010). Spots were scored as positive and retained if the signal-to-noise ratio [SNR = (signal mean – background mean)/background standard deviation] was ≥ 2.0 , the coefficient of variation (CV) of the background was < 0.8 , and the signal intensity was >200 .

Time Series Data Analysis

Time-series analysis was performed using vector autoregressive models (VAR package, Pfaff, 2007) and Granger-Causality in order to determine possible temporally

lagged relationships between geochemistry and microbial diversity. Data was first imputed for every daily time-point using the method of Forsythe et al. (1977). Stationarity of detrended data was ensured via the Augmented Fuller Dickey Test (ADF test) in R using the “tseries” package (Trapletti et al., 2018). A starting lag time (p) of p=1 was used based on historic hydrologic data (Watson et al., 2004). The lag time of p=1 represents a 2-day lag, as the data points represent every-other-day sampling. This starting lag time fell within the range of biologically relevant lag times based on calculated residence times for each well (Supplemental Table 2). Partial auto-correlation plots of the residuals were used to assess the need and fit of the VAR models (Clarke and Mirza, 2004). To determine significant relationships, Granger Causality was used in R with the “aod” package (Lesnoff and Lancelot, 2012).

Mechanisms Underlying Community Assembly

The relative roles of community assembly processes were quantified using a null-model-based framework has been proposed (Stegen et al., 2013, Stegen et al., 2015). The influence of selection was estimated based on nearest taxon index between communities (β NTI). A turnover between two communities was considered to be governed by “variable” or “homogeneous selection” when the phylogenetic dissimilarity (β MNTD, β -mean nearest taxon distance) was significantly higher (β NTI>2) or lower (β NTI<-2) than null expectation (Fine and Kembel, 2011, Webb et al., 2008; Stegen et al., 2015). Variable selection is described as the result of biotic or abiotic pressures that select for distantly related taxa from sample to sample, while homogenous selection is described as the result of biotic or abiotic pressures that select for closely related taxa from sample to sample. The

turnovers with $-2 < \beta\text{NTI} < 2$ are not differentiable from random patterns, representing the influence of stochastic processes such as neutral dispersal, drift, etc. (Stegen et al., 2013). The relative role of a process in a well was measured as the percentage of temporal turnovers governed by this process (Stegen et al., 2013, King et al., 2017). The above null model analysis was performed on the rarefied data set.

Results

Geochemical Behavior is Unique to Each Well and is Unstable Over Observed Time Period

Although the three field wells chosen for this study were in close proximity to each other, had the same underlying geology, and tapped into the same undisturbed, pristine aquifer, all three wells showed varying degrees of geochemical differences. Each well differed significantly from each other ($p < 0.01$) in several geochemical measurements (Fig. 2.1). Additionally, FW301 differed significantly from the other two wells in 7 of the 12 total geochemical parameters measured (Fig. 1, Appendix A, Supplementary Fig. A.2a-d).

FW301 showed the greatest degree of within-well variability in measured geochemistry over time, specifically for DO, conductivity, and nitrate (Fig. 2.1, Table 2.1, Appendix A Supplementary Fig. A.3). Although FW303 had the highest levels of conductivity, pH, and nitrate, it was also the most geochemical stable of the three wells over time (Fig. 2.1). Additionally, FW303 had the lowest measured concentrations of dissolved oxygen, although temporary increases were observed during weeks 7 and 8 (Appendix A, Supplementary Fig. A.3). FW305, a well that was cored within days prior to the start of this study, was more geochemically stable than FW301, and behaved uniquely

in that there was a steady increase in dissolved oxygen over time, which coincided with greater pH instability towards the later weeks of the sampled time-period (Appendix A, Supplementary Fig. A.3). Measured groundwater geochemistry over the three-month time period was similar in all three wells for oxidation-reduction potential as well as for organic acids (acetate and propionate) (Appendix A, Supplementary Fig. A.2e-g).

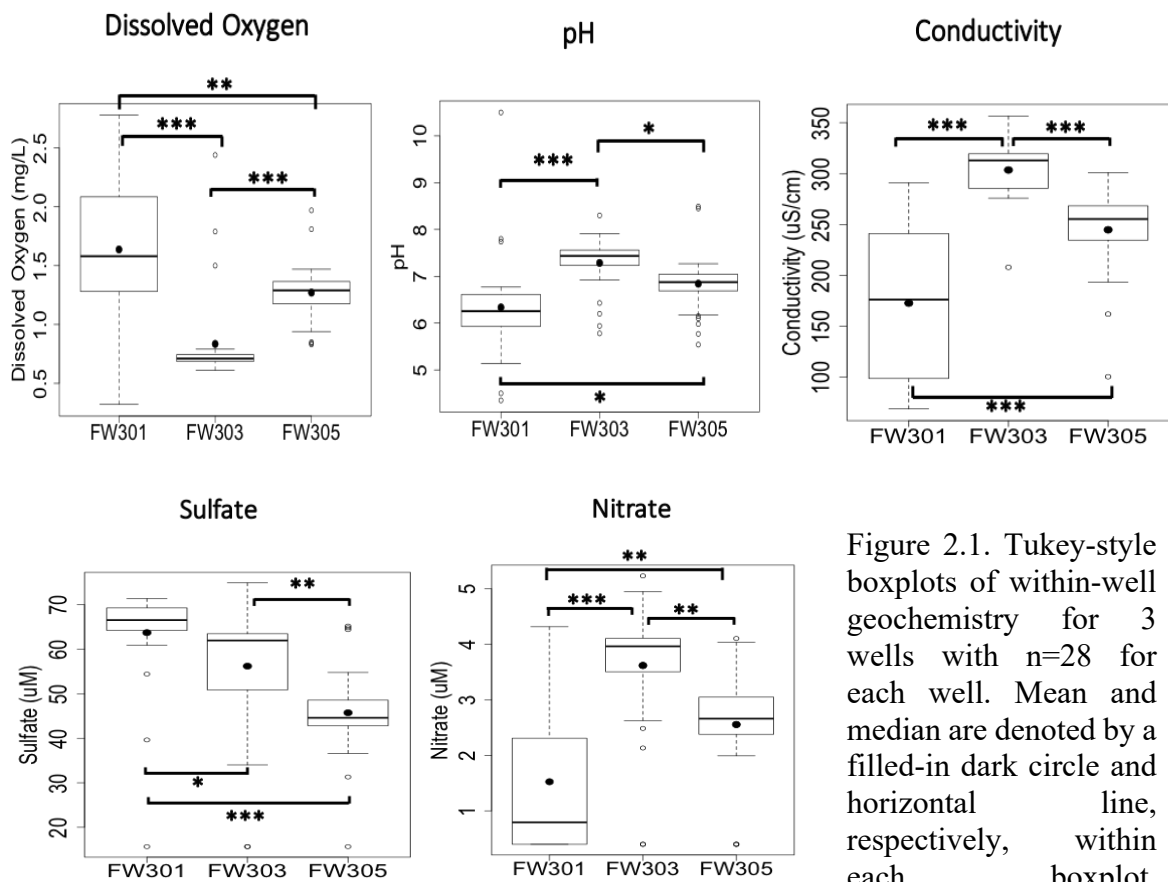


Figure 2.1. Tukey-style boxplots of within-well geochemistry for 3 wells with $n=28$ for each well. Mean and median are denoted by a filled-in dark circle and horizontal line, respectively, within each boxplot. Significance testing was

carried out via a one-way ANOVA (DF 2). Analysis of variance showed that wells differed significantly in dissolved oxygen ($F(2,81) = 25.49, p = 2.59e-09$), pH ($F(2,81) = 9.814, p = 0.000153$), conductivity ($F(2,81) = 51.05, p = 4.52e-15$), sulfate ($F(2,81) = 16.18, p = 1.23e-06$), and nitrate ($F(2,80) = 24.82, p = 4.12e-09$). Post-hoc analysis (pairwise t-tests with holm adjustment) were used to differentiate differences between groups. All three wells differed significantly from each other. Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 2.1. Boxplot statistics of measured groundwater geochemistry that displayed statistically significantly different values per well over 12 weeks

	Dissolved Oxygen (mg/L)			pH			Conductivity ($\mu\text{S}/\text{cm}$)			Sulfate (μM)			Nitrate (μM)		
	FW301	FW303	FW305	FW301	FW303	FW305	FW301	FW303	FW305	FW301	FW303	FW305	FW301	FW303	FW305
Min	0.32	0.65	0.83	5.14	6.92	5.54	73.74	275.80	213.80	60.92	34.02	36.63	0.40	2.63	1.99
L. quartile	1.28	0.70	1.13	5.86	7.08	6.38	107.40	284.40	237.00	64.24	50.88	42.82	0.40	3.51	2.38
Median	1.57	0.72	1.30	6.25	7.38	6.86	177.20	315.90	264.30	66.57	61.98	44.60	0.79	3.97	2.66
Mean $\pm$	1.69 $\pm$	0.83 $\pm$	1.27 $\pm$	6.35 $\pm$	7.28 $\pm$	6.84 $\pm$	172.76 $\pm$	303.45 $\pm$	244.75 $\pm$	66.48 $\pm$	57.06 $\pm$	45.48 $\pm$	1.64 $\pm$	3.83 $\pm$	2.82 $\pm$
SD	0.5	0.4	0.3	1.1	0.6	0.6	68.7	26.7	40.2	11.3	14.2	9.7	1.3	1.1	0.91
U. quartile	2.16	0.75	1.39	6.60	7.50	6.96	234.40	321.50	271.75	69.30	63.50	48.57	2.31	4.11	3.05
Max	2.78	0.79	1.47	6.77	7.91	7.27	291.00	356.40	300.90	71.40	74.92	54.79	4.32	4.95	4.04

Microbial Diversity and Taxonomic Composition Is Not Similar Across Wells of Close Proximity in the Same Aquifer

A total of 4,091 groundwater OTUs were identified in all groundwater samples across all three wells over the tested time period. Of the total OTUs that passed quality filtering, 1,497 (~37%) OTUs were classified as unidentifiable Bacteria (could not be identified at >80% confidence at the phylum level). The total sampled groundwater diversity included 32 phyla, 59 classes, and 153 families. Diversity profiles based on Hill Numbers, also known as Effective Species Numbers (n_{eff}), showed that overall OTU richness and unevenness was significantly higher for FW301 (p-value=0.001) (Fig. 2.2A). These trends were unaffected by sampling depth, as analysis using rarified samples show similar diversity profiles (data not shown). In addition, FW301 had a greater interquartile range at $q=1$ and $q=2$ compared to the other wells, suggesting increased temporal variability in FW-301 with respect to abundant and predominant OTUs (Fig. 2.2A).

Based on relative abundance charts, all wells are largely dominated by Proteobacteria, with classes α -, β -, and γ -Proteobacteria having the highest representation (Appendix A, Supplementary Fig. A.4). At higher taxonomic resolution, each well is dominated by unique bacterial groups, with the exception of *Comamonadaceae*, which is

highly represented in all three wells. FW303 is largely predominated by populations of Burkholderiales (Appendix A, Supplementary Fig. A.4), and FW305 contains large representations of *Oxalobacteraceae* and *Rhodobacteraceae*. In agreement with diversity measures, FW301 is more highly diverse and does not have a clearly dominating family, instead being intermittantly predominated by populations of *Comamonadaceae* and various other proteobacteria over time (Appendix A, Supplementary Fig. A.4).

A)

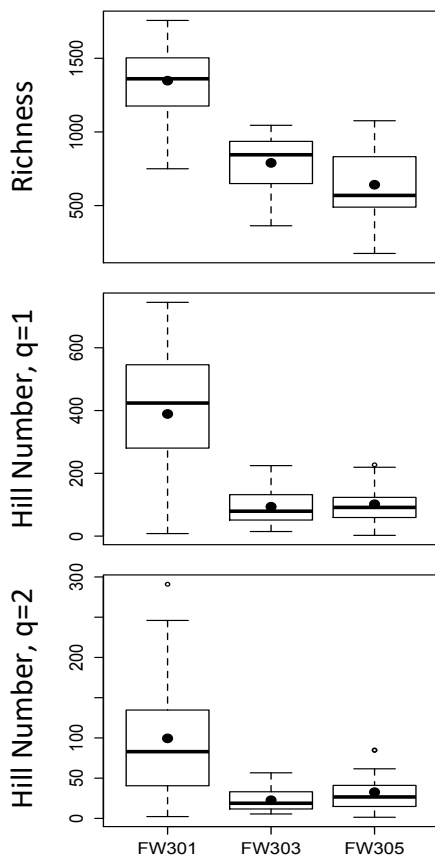
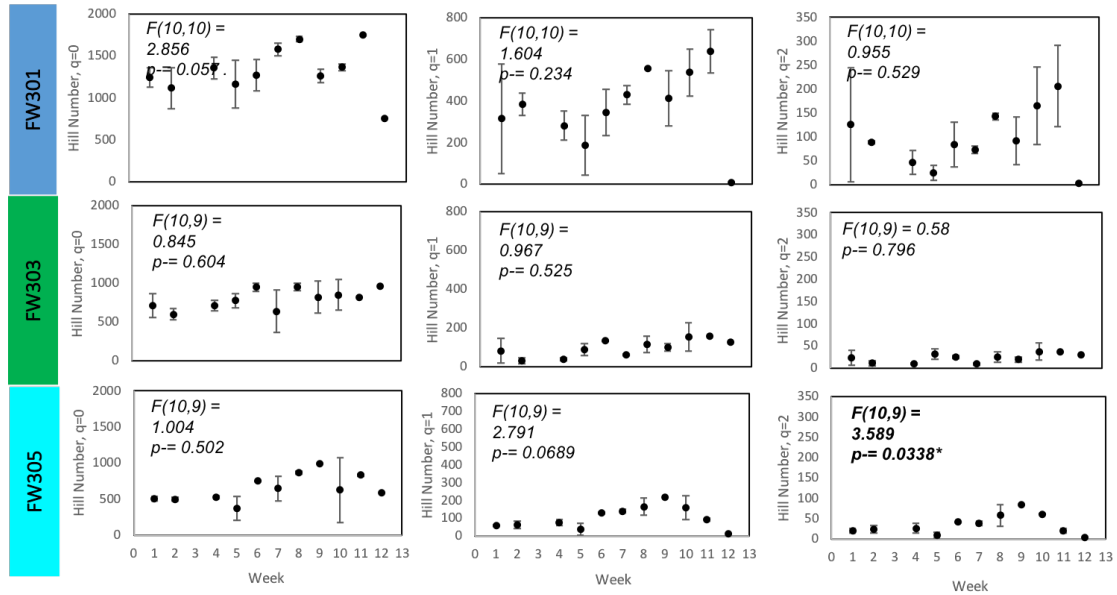


Figure 2.2. A) Tukey-style boxplots of measured within-well groundwater diversity based on Hill Numbers for $q=0$ (Richness), $q=1$ (Exponential Shannon) and $q=2$ (Inverse Simpson's). Mean (horizontal line) and median (filled-in dark circle) are denoted in each boxplot. Significance testing was carried out via a one-way ANOVA ($N=61$). Post-hoc analysis (Tukey's Honest Significant Difference) was performed to determine differences between groups. FW301 had greater OTU richness, Exponential Shannon, and Inverse Simpson's than FW303 ($F(2,54) = 43.3, p=6.0e-12$; ($F(2,54) = 33.14, p=4.1e-10$; ($F(2,54) = 13.69, p=0.000347$, respectively), and FW305 ($F(2,54) = 43.3, p=6.0e-12$; ($F(2,54) = 33.14, p=4.1e-10$; ($F(2,54) = 13.69, p=0.000332$, respectively). FW301 also has a greater interquartile range than FW303 and FW305 for Exponential Shannon (IQR = 259.2; 83.9; 64.1, respectively) and Inverse Simpson's (IQR = 95; 21.9; 26.3, respectively) but not richness (IQR = 307; 275; 324.5, respectively). This indicates higher temporal variability in FW301 for measured abundant and dominant OTUs, but not for measured richness. Results are similar for rarefied samples. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

B) Plots show weekly means and standard error bars (SEM) for within-well groundwater diversity. Data from 2 days per week ($n=2$) were analyzed, with the exception of weeks 9 and 11 where only one date was sampled in wells FW305 and FW303, respectively, and week 12 were only one date was sampled in all three wells. Week 3 shows no weekly mean because no samples were collected that week. Test of significance was carried out via a one-way ANOVA in R on all samples taken over 11 weeks for a total of 21 samples in well FW301, and 20 samples in the other two wells.

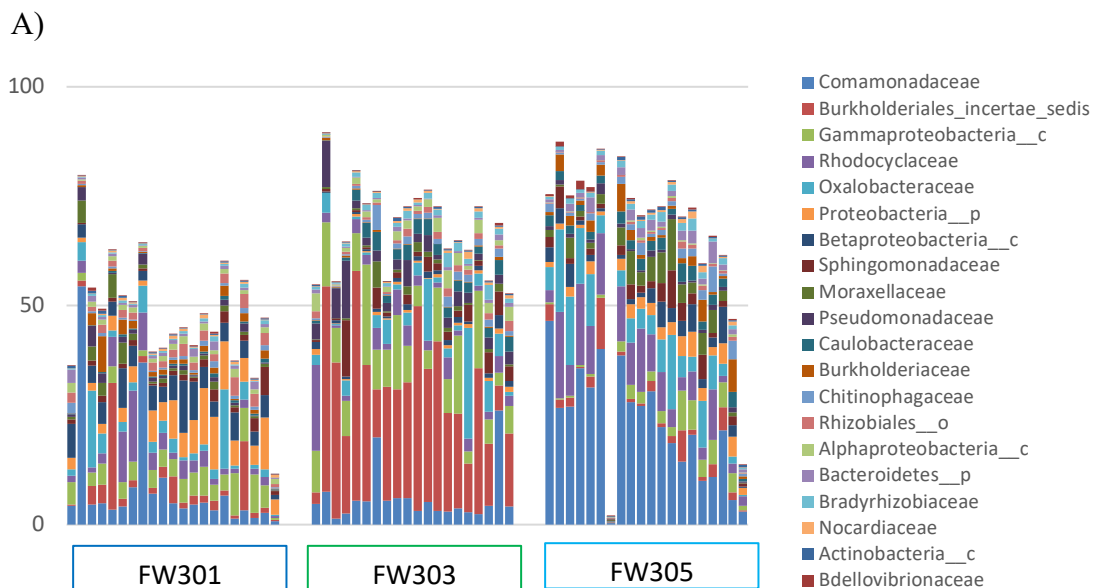
B)



Microbial Community Structure is Unique Over Time Across Wells

Weekly mean richness varies significantly in FW301 ($F(10,10)=2.856$, $p=0.057$) but not in FW303 ($F(10,9)=0.845$, $p=0.604$) or FW305 ($F(10,9)=1.004$, $p=0.502$), indicating that rare and transient populations contribute to much of the variability measured in FW301 but not in the other two wells (Fig. 2.2). As q increases (effects of rare OTUs are removed), mean weekly diversity differs significantly in FW305 ($F(10,9)=2.791$, $p=0.0689$ for $q=1$ and $F(10,9)=3.589$, $p=0.0338$ for $q=2$) but not for FW301 or FW303. This indicates weekly differences in community structure based on both abundant and dominant communities over time in FW305 but not in the other two wells. These same results are obtained when performed on rarified samples. Standard mean errors (SEM) are high for FW301. Although ANOVA's did not show significant difference across weeks as q increased, high SEM indicates high within-week variability in FW301.

Only 20 bacterial taxa were observed in all groundwater samples across all three wells (Fig. 2.3, Table 2.2). As such, these 20 taxa can be considered cosmopolitan in the system over the tested time period. However, 14 of these 20 taxa were observed at low relative abundances in each well (<1%) on most sampled dates (Table 2.2; Fig. 2.4). In other words, only six of the cosmopolitan taxa (*Comamonadaceae*, Burkholderiales *incertae sedis*, unclassified γ -Proteobacteria, *Rhodocyclaceae*, *Oxalobacteraceae*, and *Sphingomonadaceae*) were moderately to highly abundant (>1%) in the majority of samples in each well. Additionally, 18 non-cosmopolitan taxa were observed to be highly abundant (>5%) in at least one groundwater sample over the tested time period (Fig. 2.3B). At the phylum level, each well experienced periodic blooms of Firmicutes and *Deinococcus-Thermus*; however, the dates and intensities of these blooms



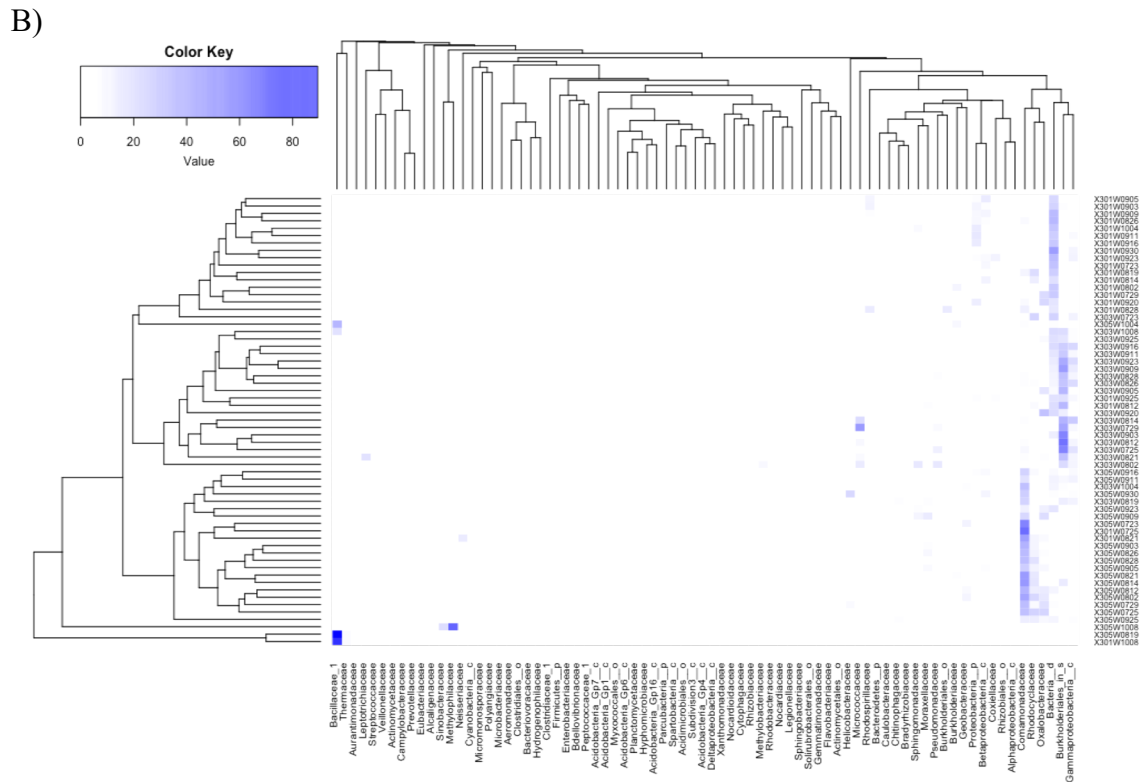


Figure 2.3 A) 20 cosmopolitan taxa found in all groundwater samples. Out of 218 total unique taxa identified, only 20 were found in all groundwater samples. (~1%). These 20 taxa account for 1.9-89.6% of the daily measured groundwater diversity. All 20 taxa were found in >1% abundance in at least one groundwater sample over time. B) Transiently abundant taxa found in >1% in at least one groundwater sample. 18 of these taxa are found at >5% relative abundance in at least one sampled time point.

Table 2. Number of times cosmopolitan family appeared at >1% over 3-month sampling period. On most days, cosmopolitan taxa were lowly abundant in groundwater communities. Only 6 cosmopolitan taxa appeared at >1% on more than half of the days sampled in all wells (bold).

Family	N=21 FW301	N=20 FW303	N=20 FW305
Comamonadaceae	20	20	19
Burkholderiales_incertae_sedis	18	20	12
Gammaproteobacteria__c	20	20	13
Rhodocyclaceae	20	19	19
Oxalobacteraceae	17	16	18
Proteobacteria__p	21	3	17
Betaproteobacteria__c	20	8	16
Sphingomonadaceae	12	14	16
Moraxellaceae	9	5	16
Pseudomonadaceae	9	20	4
Caulobacteraceae	0	16	18
Burkholderiaceae	11	0	10
Chitinophagaceae	5	13	11
Rhizobiales__o	18	10	1
Alphaproteobacteria__c	14	15	0
Bacteroidetes__p	5	2	15
Bradyrhizobiaceae	1	12	13
Nocardiaceae	0	1	2
Actinobacteria__c	0	0	0
Bdellovibrionaceae	0	0	2

were unique in each well. These results suggest that while only 6 taxa were observed to consistently predominate groundwater samples, an additional 18 taxa transiently predominated over time in groundwater. The high level of variable representation across all wells contributes to a high degree of dissimilarity in microbial populations in each well. The wells are between 56-79% dissimilar to each other with regard to the OTUs present at any given time in any given well, and on average about 80% dissimilar with regard to the spatial and temporal shifts in the abundances of shared OTUs (Appendix A, Supplementary Fig. A.5).

While the term “cosmopolitan” was meant to designate taxa that were found present across all sampled background wells, the degree of stable residency versus transience

within individual each well was also investigated (Supplementary Figures A7-A9, Supplementary Table A3). Within each well, a core-set of OTUs were identified as “resident” if they were present in every sampled time point for that well. The degree of transience was determined based on the percentage of all samples for which that OTU was not observed/missing/absent from the dataset of that well. While each well had a unique set of core, or resident, OTUs, sampled groundwater from FW301 and FW305 appeared to share more resident OTUs over the sampled time period (Supplementary Figure A10).

Microbial Community Structure is Highly Variable Within Each Well Over Time

Within-well variability of OTUs present and the proportional representation in the respective community were observed to be very different from day to day, and the degree of temporal variability within a single well was largely unexpected. Between sampling points, the OTUs present in any given well differed on average between 41-74% (Appendix A, Supplementary Fig. A.6, Supplementary Table A.1), which contributed to significant differences in within-well diversity from week to week (Fig. 2.2B). While the 20 shared cosmopolitan taxa represented a level of stability in the system, temporal relative abundances of these taxa changed uniquely in each well (Fig. 2.2A). Daily cumulative relative abundances of the 20 cosmopolitan taxa fluctuated over time, ranging from 1.9 to 89% of the total daily measured groundwater diversity (Fig. 2.3, Appendix A, Supplementary Fig. A.6). However, in FW303, the 20 cosmopolitan taxa maintained a cumulative relative abundance of 50% or greater in all samples. Therefore, in FW-303, cosmopolitan taxa made up over half of the community on all sampled time-points, indicating a level of structural stability not observed in the other two wells.

SIMPER analysis was used to calculate the Bray-Curtis dissimilarities between samples based on specific populations within each sample, thus calculating the contributions of specific populations to the weekly changes in community structure in each well. In FW-301, OTUs mostly classified as β -Proteobacteria (*e.g.*, Burkholderiales, *Comamonadaceae*, *Rhodocyclaceae*) contributed significantly to the weekly differences (Appendix A, Supplementary Fig. A.6, Supplementary Table A.1). However, FW301 was unique in that the cumulative abundance of contributing OTUs was typically less than 40% of the total measured abundance for any given date. This data indicates that the majority of OTUs contributing to weekly dissimilarities in FW301 were moderately to rarely abundant on any given day, with periodic blooms of different OTUs contributing largely to weekly dissimilarities. In FW303, changes in relative abundances of the consistently present Burkholderiales *incertae sedis* contributed highly to the weekly differences. Additionally, unlike in FW301, the majority of OTUs with the most contributions to weekly dissimilarities were often highly abundant. This indicates that the differences in this well were attributed to changes in the consistently abundant, as opposed to the rarely transient, populations. FW305 had the highest proportion of *Comamonadaceae* populations; however, a steady decline in relative abundances can be observed over time (Fig. 2.3A, 2.4, Appendix A, Supplementary Fig A.6). *Rhodocyclaceae* and *Oxalobacteraceae* also contributed significantly to changes in community structure over time in FW-305 (Fig. 2.4 and Supplementary Fig A.6).

An unconstrained ordination revealed patterns in FW-305 not observed in the other two wells, and distinct shifts in the populations were observed that separated the sampled

communities by time of sampling (Fig. 2.5A). Samples collected in July/August were clustered, followed by a sudden shift in the community in mid-August. Communities sampled in the late August and September cluster distinctly apart from those of the earlier months. The last two time points, sampled in October, are dramatically different in overall composition than any of the samples. Interestingly, FW-301 and FW-303 were cored approximately 20 years ago, while FW-305 was cored a few weeks prior to the start of this study. The unconstrained analysis may reveal unique successional dynamics occurring as the surrounding subsurface environment was disturbed during the coring and the bacterial community responds to altered flow and/or geochemical conditions.

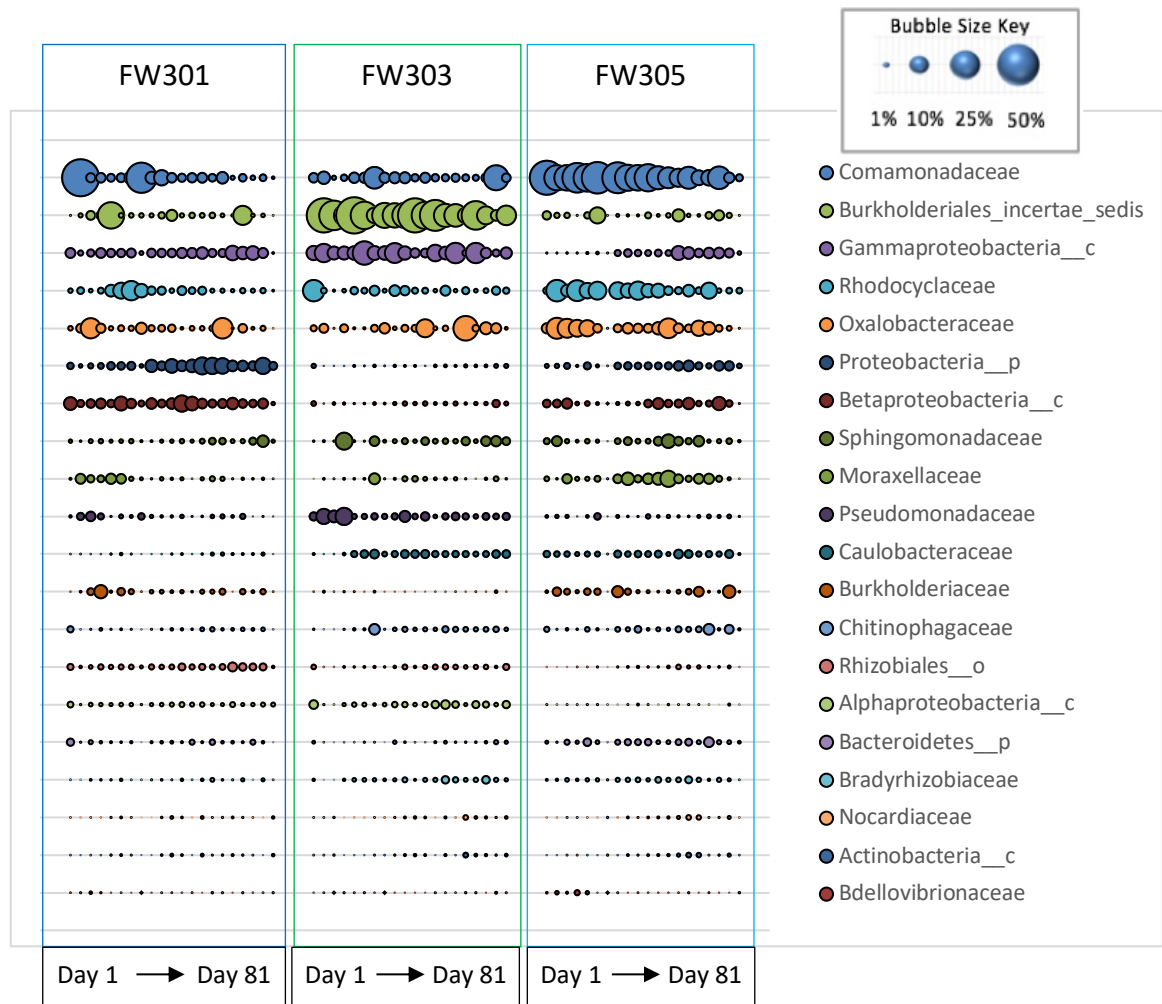
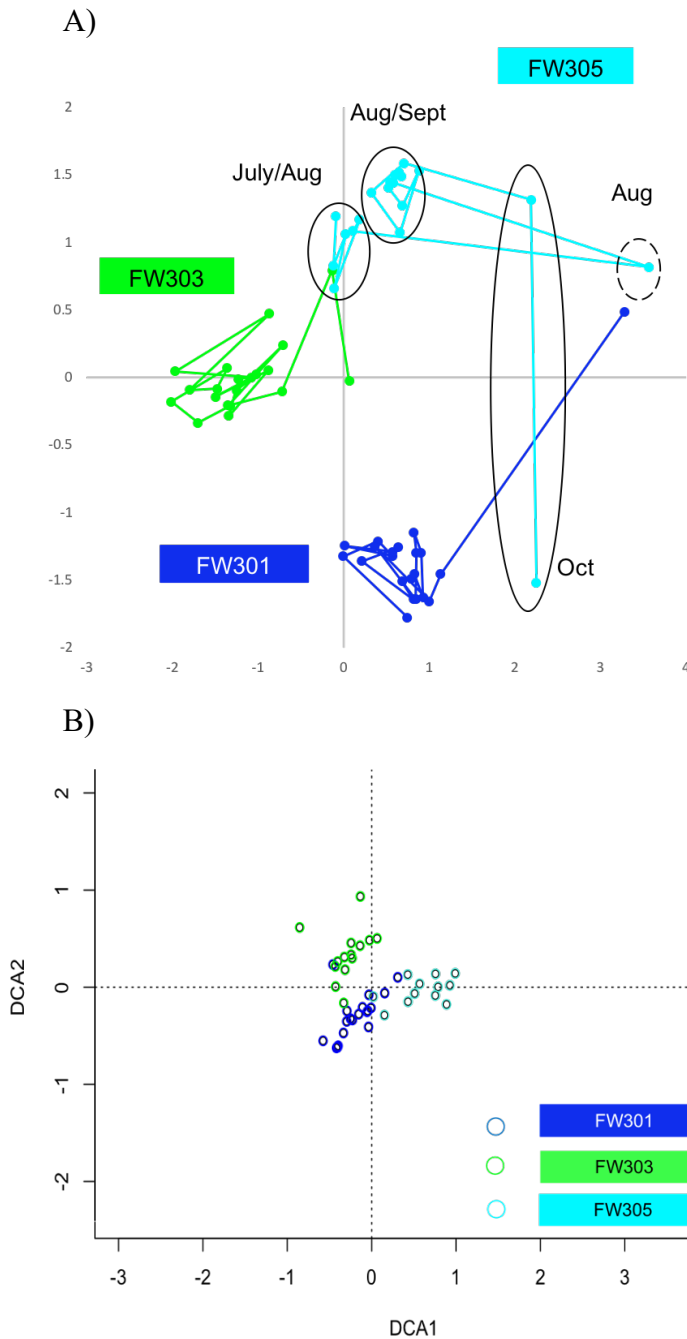


Figure. 2.4 Bubble plot of relative abundances at the highest classifiable level for cosmopolitan groups in sampled groundwater. The timescale for samples begins on Day 1 of field sampling (July) and ends on Day 81 (October) for each well grouping. Bubbles size represents relative abundance within sample and were scaled to 25% of original bubble size to reduce overlaps between bubbles.



Functional Gene Diversity Displays Less Spatial and Temporal Dissimilarity

Based on functional gene arrays, the functional gene diversity of the groundwater communities in the tested wells was more similar over time than the phylogenetic diversity. (Fig. 2.6) Additionally, the intra-well variability was also much less compared to the SSU rDNA gene sequences. The functional potential within and across samples remained relatively constant throughout the three-month study (Fig. 2.6 and Table 2.3). All three wells were enriched in COGs classified as metal homeostasis (~27%), stress (~16%), carbon cycling (~16%) and virulence (~15%).

Changes in Diversity Relate to Geochemical Fluctuations Uniquely in Each Well

Canonical correlation analysis (CCA) indicated potential geochemical drivers of community composition in each given well (Fig. 2.7). According to the CCA plots, changes in fluoride, conductivity, sulfate, and DO are the greatest drivers for differences in community composition between the three background wells. One limitation of correlation analysis is that it might not capture potential delays in response between changes in geochemistry and diversity within a well, and vice versa. Additionally, time-series data are often inherently dependent and therefore do not meet the requirement of statistical independence (Pfaff, 2008).

To account for potential effect of time delayed responses as well as issues with dependent data, vector autoregressive models (VAR) were used (Fig. 2.8, Table 2.4). The results of the VAR models indicate that both dissolved oxygen and chloride are significantly affected by increases in richness in FW301. In the case of dissolved oxygen,

A)

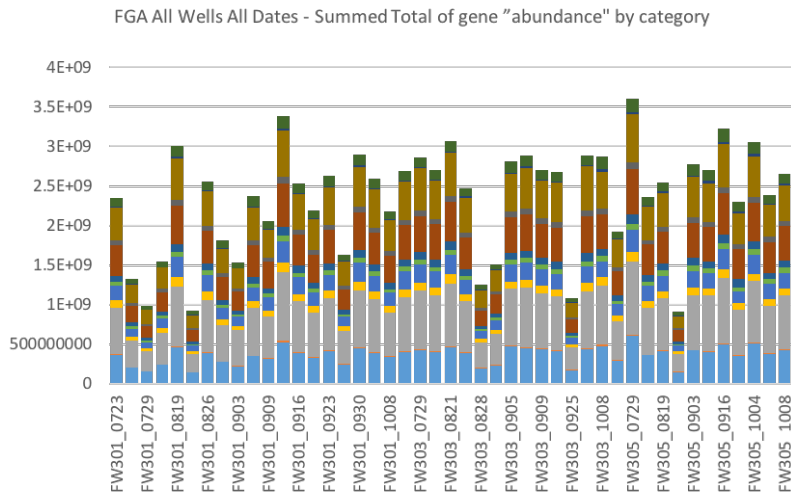
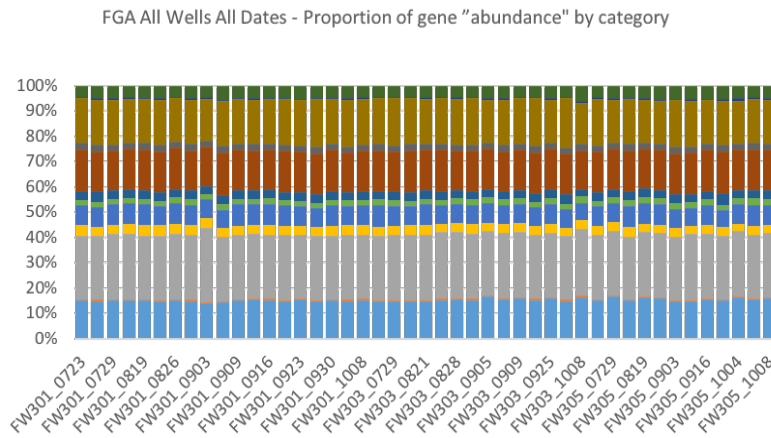


Figure 2.6 Functional Gene Array (GeoChip). A) Proportion of individual genes (by category) remain consistent within and across wells. B) Summed Total of gene signal abundance by category. C) Pie charts representing total proportion of COG's by category

B)



C)

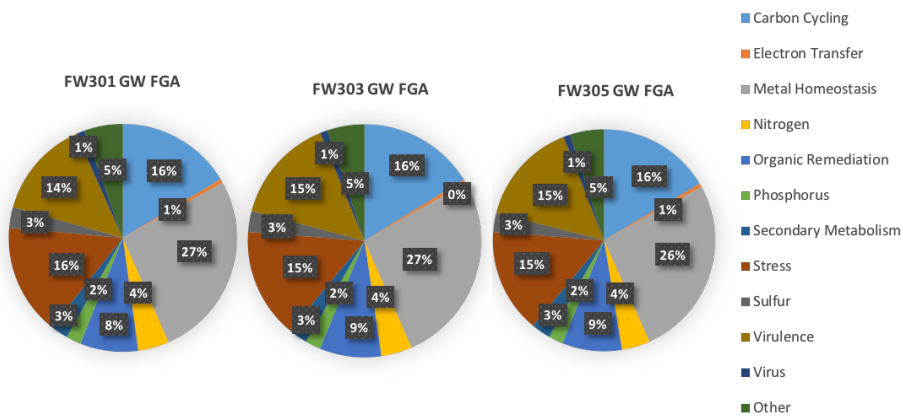


Table 2.3. Number of COG's and potential function

Table 3. Number of COG's and potential function			
	# of enriched COGs	# of enriched COGs	# of enriched COGs
Potential function	FW301	FW303	FW305
Carbon Cycling	14851	10629	8126
Electron Transfer	498	349	256
Metal Homeostasis	23990	17478	12954
Nitrogen	3978	2800	2034
Organic Remediation	7449	5598	4277
Phosphorus	1934	1403	1054
Secondary Metabolism	2268	1739	1388
Stress	14226	9968	7467
Sulfur	2681	1914	1378
Virulence	12635	9621	7436
Virus	1106	677	487
Other	4946	3381	2434
Total	90562	65557	49291

increases in richness corresponded to decreases DO concentrations after 6 days (*p*-value, 0.035). On the other hand, increases in richness were accompanied by increases in chloride with a delay of 2 days (*p*-value, 0.022). For FW303, ORP and conductivity had significant relationships with changes in richness. Increases in richness were observed 4 days after increases in conductivity (*p*-value, 0.00195). Interestingly, while the CCA plot showed that ORP only had a minor influence in driving community composition in any well, the VAR model indicated a strong bi-directional Granger Causality between ORP and richness following a 4-day time-delay (Table 2.4).

Variables are considered to be cointegrated when long-run equilibria are displayed (Masih and Masih, 1996). Relationships in FW305 showed a high degree of cointegration

(data not shown) such that traditional time-series analysis using VAR models would have resulted in spurious results (Pfaff, 2008). All modeled variables for FW305 were cointegrated with richness, with the exception of the chloride and fluoride anions. Future work includes development and analyses with more appropriate models (*e.g.*, Vector Error Correction Models) for the FW305 dataset.

Different Ratios of Deterministic and Stochastic Forces

In sampled groundwater of the tested wells, deterministic selection was the dominant calculated process, with the majority of β NTI comparisons being <-2 , suggesting strong biotic or abiotic pressures worked to select for closely related taxa from one day to the next (Table 2.5). Based on the β NTI analysis, each well was also at least partially influenced by both deterministic and stochastic processes, although each well was influenced by either process in different proportions (Fig. 2.9). Of all wells, FW301 was the most affected by stochastic processes (35%), which agrees with the high temporal variability observed in this well. FW305 was primarily governed by strong selective pressures (93% homogeneous selection) and was the well that was least influenced by undominated (stochastic) processes (7%). This result supports the idea that this newly constructed well is undergoing deterministic selective pressures to shape microbial community composition.

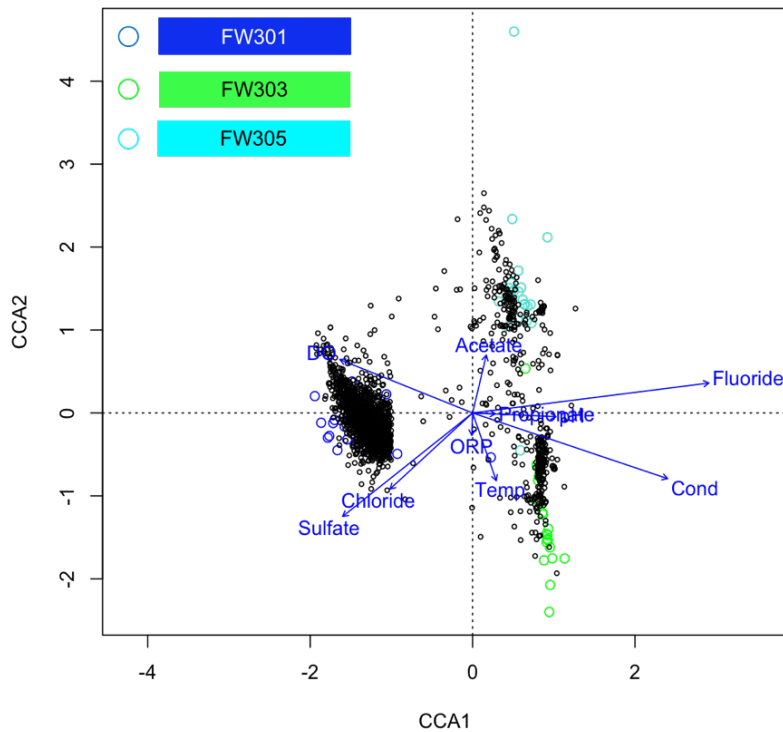


Figure 2.7. Multivariate ordination showing associations between groundwater community structure and geochemical variables using Canonical Correspondence Analysis (CCA). Small black dots represent individual OTUs, larger colored circles represent samples for each well, and blue vectors represent association between communities and geochemistry. Strength of association is inferred from vector length.

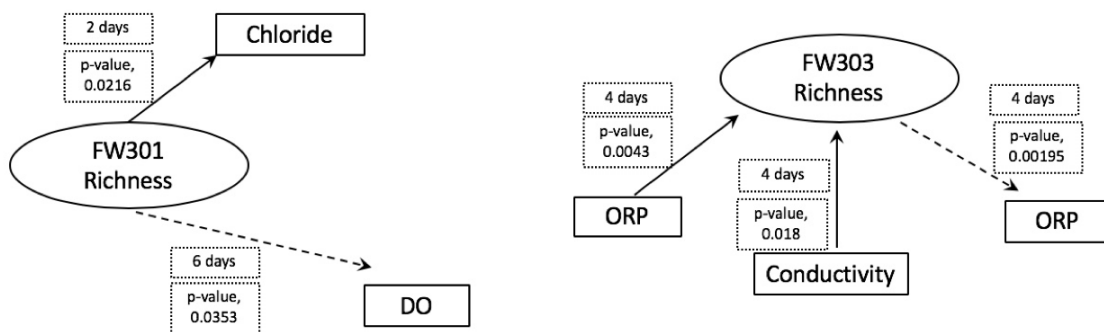


Figure 2.8. Association networks based on results of Vector Autoregressive Models of diversity based on Hill Numbers and geochemical relationships. Significant lagged relationships ($p\text{-value} \leq 0.05$) and the lag value in days are shown. Positive associations are represented by solid edges, while negative associations are represented by dashed edges. The direction of the arrow indicates the direction of the delayed relationships - the node preceding the arrow is ahead of the node which follows the arrow.

Table 4. Results of VAR Model and Granger Causality using the Wald Test for a given lag (p) chosen that eliminated autocorrelation. NA For FW-303, the VAR model could not be performed because the data was non-stationary.

Granger Causality (p-values) for Richness and Geochemistry in FW301			
Richness "Granger Causes":	p-value	Richness is "Granger Caused":	p-value
Temp	0.079	Temp	0.5
Cond	0.36	Cond	0.26
ORP	0.065	ORP	0.37
Nitrate	0.78	Nitrate	0.098
Sulfate ⁺	0.013	Sulfate	0.99
pH	0.5	pH	0.11
DO*	0.00026	DO	0.088
Fluoride	0.58	Fluoride	0.16
Chloride*	0.036	Chloride	0.99
PRCP	0.9	PRCP	0.65

Granger Causality (p-values) for Richness and Geochemistry in FW303			
Richness "Granger Causes":	p-value	Richness is "Granger Caused":	p-value
Temp	0.64	Temp	0.76
Cond	0.18	Cond*	0.016
ORP*	0.0014	ORP*	0.0059
Nitrate	0.095	Nitrate	0.96
Sulfate	0.063	Sulfate	0.93
pH	0.72	pH	0.62
DO ⁺⁺		DO ⁺⁺	
Fluoride	0.64	Fluoride	0.96
Chloride	0.4	Chloride	0.61
PRCP	0.98	PRCP	0.78

Granger Causality (p-values) for Richness and Geochemistry in FW305			
Richness "Granger Causes":	p-value	Richness is "Granger Caused":	p-value
Temp ⁺⁺		Temp ⁺⁺	
Cond ⁺⁺		Cond ⁺⁺	
ORP ⁺⁺		ORP ⁺⁺	
Nitrate ⁺⁺		Nitrate ⁺⁺	
Sulfate ⁺⁺		Sulfate ⁺⁺	
pH ⁺⁺		pH ⁺⁺	
DO ⁺⁺		DO ⁺⁺	
Fluoride	0.84	Fluoride	0.54
Chloride	0.82	Chloride	0.68
PRCP ⁺⁺		PRCP ⁺⁺	

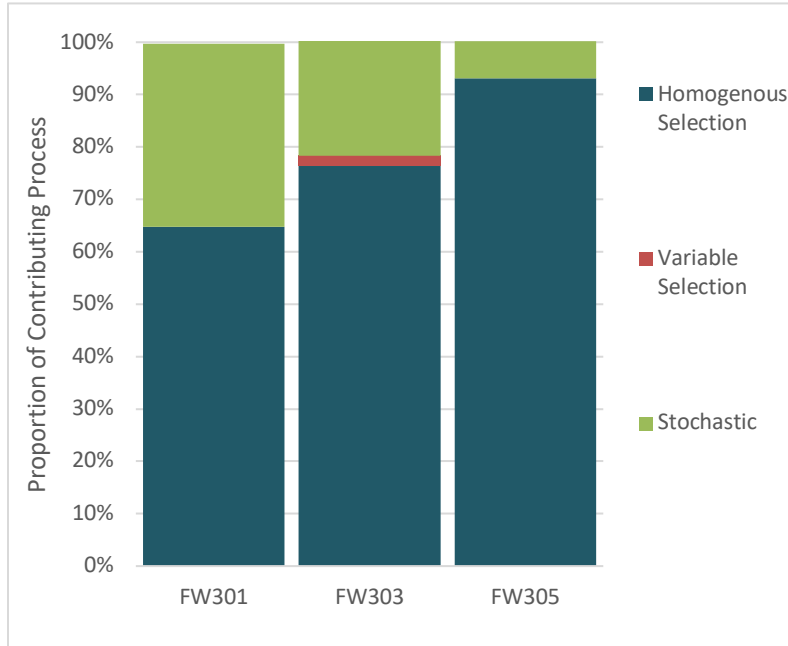


Figure 2.9. Relative influence of community assembly processes on temporal turnover of within-well bacterial assemblages. Processes were calculated using RC_{Bray} and $bNTI$ as described in Materials and Methods.

Discussion

Groundwater systems and the indigenous microorganisms play a vital role in sustaining life, both above and below the subsurface through major contributions to the water cycle (Goldscheider et al., 2006). Understanding the factors that affect microbial community structure and function is critical to meet present and future global demands of water maintenance and utilization. However, the connection between microbial diversity, biogeochemical processes, and the flow of energy and resources between shallow subsurface ecotones is poorly understood. In addition, shallow subsurface ecotones include both groundwater and porous medium, and therefore, both aqueous and solid media are important and unique niches that impact resource allocation. The current study focused on the variability of the bulk aqueous phase of an uncontaminated, oligotrophic aquifer.

Despite having hydrological, chemical, and geological heterogeneity, most groundwater systems are thought to be constant and predictable due to the nature of oligotrophic environments through similar transition zones (Griebler and Lueders, 2009). Based upon these common sentiments, temporal hydrological events have been hypothesized to impact geochemical changes, particularly via redox shifts caused by hydraulic events over the time span of seasons (Vroblesky and Chapelle, 1994; Haack et al., 2004; Zhou et al., 2012). More recently, temporal dynamics have also been attributed to environmental disturbances, for example nutrient influxes (*e.g.*, hot spots and hot moments) at the interface between terrestrial and aquatic ecosystems (McClain et al., 2003), or in the case of water systems, the influx of additional water from the surrounding environment. For example, it has been shown that there can be direct transfer of nutrients and microbial populations within transitionally saturated and permanently saturated zones (Bougon et. al., 2011; Yabusaki et al., 2017). Studies of the hyporheic zone along the Columbia River in Washington, USA have shown that seasonal changes in river height and overflow due to increased rainfall result in increased diversity of aquifer microbial communities (Lin et al., 2012). However, the opposite effect on diversity has also been documented. Rising groundwater levels due to seasonal effects on water table recharge were shown to accompany a drop in bacterial diversity (attributed to dilution effects) in spring and summer months in an alpine oligotrophic porous aquifer (Zhou et al., 2012).

The described study focused on the temporal dynamics of an uncontaminated groundwater system with limited hydraulic connectivity between sampling wells into groundwater housed within the same underlying geological formations. Both geochemistry

and microbial communities were uniquely dynamic over space and time with varying degrees of fluctuation intensities and sparsely shared populations. In fact, few populations could be considered cosmopolitan for the three wells or over the measured time span. The results indicated a high degree of transience in groundwater bacterial populations with the majority of measured groundwater populations (approximately 93%) being sporadically over time. Despite the high level of transience, microbial cell numbers in the described study ranged between 10^5 and 10^6 cells/ml (data not shown) and this is in general agreement with previous groundwater studies (Griebler and Lueders, 2009).

In this study, rainfall did not correlate to measured changes in diversity or geochemical fluctuations in the tested wells (data not shown). However, it is worth noting that the study was conducted during the transition period between the drier summer months of July-August, and the fall months of September-October. Given that recharge in this aquifer occurs primarily via heavy rainfall events and vertical percolation of excess water to the saturation zone (Solomon et al., 1992), which typically occur during the wetter winter months, the results indicate a high degree of variable biodiversity during the dry period. Future work is underway to better understand microbial community dynamics during and between different seasons for the uncontaminated and contaminated carbonate-rock aquifers and the implications for ecosystem function.

Temporal dynamics of microbial communities have been studied in other environments, such as the human microbiome (Koenig et al., 2011), plant-flower (Shade et al., 2013a), surface soils (Wertz et al., 2007), and other aquatic systems (Hofle et al., 1999; Shade et al., 2007). A meta-analysis surveying the temporal variability across various

biomes (Shade et al., 2013b) demonstrated that areas with high physicochemical instability (*e.g.*, stream, human palm) varied considerably over time, while more *relatively* stable environments (*e.g.*, soil) varied less considerably, suggesting that microbial communities might vary predictably across environments over time depending on physical and chemical gradients. Additionally, environments likely undergoing succession (*e.g.*, infant gut, flower surfaces) display high levels of diversity and community composition shifts over time. In the described study, groundwater from a pristine aquifer previously believed to be geochemically stable displayed temporal changes of geochemistry (*e.g.*, ORP, DO, pH, and conductivity) to varying extents over short time scales (days) that were statistically significant (Fig. 2.1). For FW301, the VAR models indicated that increases in species richness corresponded to subsequent decreases in DO levels (6 days) and increases in chloride (2 days). A possible explanation could be altered population distributions are linked to microbial activity that impacts DO consumption as well as anion/cation balances between groundwater and sediments. In addition, fine-scale changes (micro-environments) in pH associated to microbial activity could also impact anion/cation interactions with sediments. For FW303, ORP and conductivity were predicted to have bidirectional causal relationships with species richness. Similar to FW301, microbial activity could be linked to the population distributions and thereby impact ORP and/or balance of anion/cations between groundwater and sediments. Further work is needed to delineate these relationships, for example, bench-scale packed bed reactors in which these field parameters can be controlled.

It is typically assumed that low levels of biodiversity may allow a conditionally

consistent ecosystem to function efficiently while greater biodiversity would be needed in fluctuating environments (Humbert and Dorigo, 2005). Most of the work that has been done with pristine aquifers suggests more constant microbial communities with low diversity mainly attributed to a more consistently oligotrophic environment (Griebler and Lueders, 2009). However, with greater depth of spatial and temporal analyses, our results indicate that a shallow groundwater system does experience geochemical and microbial populations changes on short time frames. Although GeoChip analysis of functional diversity predicted a functionally stable community, this result could coincide with the notion of functional redundancy (Naeem and Li, 1997) in which high turnover of lowly abundant populations could contribute significantly to maintenance of overall functions (Jousset et al., 2017). Moreover, different geochemical parameters were shown to have significant relationships with changes in richness and diversity (Fig. 2.7 and 2.8, Table 2.4), but these relationships were different for each well. In fact, although homogenous selection was estimated to be the dominant ecological process for the three tested wells, each well was predicted to have a unique ratio of deterministic and stochastic selection forces according to β NTI analyses. Future work is needed to track baseline system parameters over time in order to better understand perturbation impacts (*e.g.*, extreme weather) and the impact on long-term system functions.

In the most newly constructed well (FW305), unique relationships observed suggest that this well may have been undergoing major shifts in community composition likely attributed to succession events following the coring event (Fig. 2.4). The microbial variability in FW305 occurred among the most dominant community members (SIMPER).

The DCA ordination also showed evidence of major compositional shifts over the three-month sampling period. In addition, richness in FW305 was cointegrated with 10 of the 12 geochemical parameters measured. By definition, data that are co-integrated *must* display Granger causality in at least one direction (Pfaff, 2008). In other words, evidence of co-integration between variables “rules out Granger non-causality”, implying either unidirectional or bidirectional Granger Causality (Masih and Masih, 1996). Since FW305 was only a few weeks old prior to the start of sampling, the unique relationships observed in this well likely describe the long-term inter-connected relationships between hydrology, geology, and microbial populations that became disrupted due to well construction. Further work is needed to track temporal variability of planktonic and sediment associated microbial communities in established and new wells, including sediment surrogates (King et al., 2017; Smith et al., 2018) to better understand community dynamics impacted by coring.

Interestingly, the well with the highest overall geochemical and microbial variability, FW301, also has a screen interval roughly 3x larger than FW303 and 2x larger than FW305. It is possible that the sampled formation water originated from a greater range of depths from the subsurface profile, or that impacts of periodic percolation events that occurred over the three-month period were more represented within the sampled groundwater from this well. The greater sampling depth could result in the variable representation of otherwise dispersion-limited microorganisms originating from shallower depths of the soil profile, variable infiltration of total dissolved solids from the surrounding weathered saprolite, and/or higher geochemical perturbations (*e.g.*, DO, organic).

Differences in these environmental parameters across depth corroborate the idea that changes in hydrology and groundwater chemistry impact groundwater microbial communities (Flynn et al., 2009), and future work will focus on soil sampling across the transition zones and water sampling from wells with discrete depth intervals across longer time periods.

The existence and ubiquitous nature of rare microbial taxa in all types of environments has been well documented (Sogin et al., 2006; Bent and Forney, 2008; Fuhrman et al., 2009; Pedrós-Alió, 2012), and more recent analysis of temporal studies (air, water, soil, human) have also documented the presence of rare taxa that occasionally bloom to dominant members of the community or remain at low abundance (Shade, 2014; Joussett et al., 2017). The "conditionally rare taxa" (CRT) were recently estimated to make up to 28% of community membership, represent a broad diversity of bacterial and archaeal lineages, and explain large amounts of temporal community dissimilarities (Shade, 2014). While the presence of periodically abundant populations (including cosmopolitan taxa) were observed in all three wells, FW301 had a much higher contribution of CRT to the dominant taxa than the other two wells, coinciding with higher overall diversity and geochemical instability over the sampled time period. Whether this is ecologically significant, or a consequence of screen interval length is currently unknown; however it may suggest an important role for rare taxa in the localized area of the aquifer. In contrast, sampled groundwater from a newer well, FW305, showed higher temporal variability in the dominant populations, indicating selective pressures in this recently disturbed system.

Given the unexpected extent of temporal variability in transient bacterial populations for uncontaminated groundwater based upon commonly used pyrotag data, the potential contribution of 'active' and 'inactive' cells is not known. However, the observed community dissimilarity for each well was constant over the tested time period, and the composition of the diversity changed dependent upon the well. In addition, similar predominant populations (Proteobacteria, Actinobacteria, and Bacteroidetes, Acidobacteria, Chloroflexi, and Nitrospirae) were observed as in previous groundwater studies (Griebler and Lueders, 2009). Moreover, analysis of the same samples with the Geochip demonstrated that functional gene diversity was more constant over the same time period. It should be noted that Geochip sensitivity (microarray) is different compared to direct sequence comparisons for the SSU rDNA genes and future work should include functional gene diversity for selected processes of interest and/or metagenomes. In addition, current work is underway to track general and specific microbial activity over time for groundwater systems to relate biodiversity and ecosystem function, and our preliminary results indicate that BONCAT (amino acid tagging) can be used to quantify active cells in EOR-FRC groundwater. Further investigations into seasonal variability within and across wells, as well as the effects of water table recharge on each well is warranted. Investigations into the relationship between microbial diversity, ephemeral populations, transient geochemical flux, and the contributions to overall microbial activity in groundwater systems are currently underway.

Conclusions

Freshwater resources are becoming jeopardized at alarming rates. Full assessment of undisturbed groundwater composition, both biotic and abiotic components, over space and time are necessary to establish expected variability over both short- and long-term scales. Without such baselines, observed changes to groundwater composition could be erroneously attributed to anthropogenic effects or other disturbance events. This study showed for the first time that groundwater composition of an undisturbed aquifer varies dynamically and uniquely over short time scales.

Acknowledgments

This material is funded by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (<http://enigma.lbl.gov>), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231.

CHAPTER THREE

POPULATION FILTERING IN SEDIMENT BIOFILMS FROM DYNAMIC, SOURCE
PLANKTONIC COMMUNITIES

Contribution of Authors and Co-Authors

Manuscript in Chapter Three

Author: Anna J. Zelaya

Contributions: Performed literature review, synthesized ideas, performed statistical and bioinformatical analyses, wrote and edited the manuscript

Co-Author: Albert E. Parker

Contributions: Statistical analyses, reviewed and edited the manuscript

Co-Author: Kathryn L. Bailey

Contributions: Performed field samplings, geochemical analysis and data collection, reviewed and edited the manuscript

Co-Author: Ping Zhang

Contributions: Performed DNA extraction and paired-end sequencing on all samples, reviewed and edited the manuscript

Co-Author: Joy Van Nostrand

Contributions: Performed DNA extractions, paired-end sequencing, and functional gene array, wrote methods, reviewed and edited the manuscript

Co-Author: Ning Daliang

Contributions: Performed β MNTD and β NTI analysis, wrote methods, analysis reviewed and edited the manuscript

Co-Author: Dwayne A. Elias

Contributions: Supervisor and Principal Investigator to Kathryn L. Bailey, reviewed and edited manuscript

Co-Author: Jizhong Zhou

Contributions: Supervisor and Principal Investigator to Daliang Zhou, reviewed and edited manuscript

Co-Author: Terry Hazen

Contributions: Synthesized ideas, reviewed and edited the manuscript

Co-Author: Adam P. Arkin

Contributions: Synthesized ideas, reviewed and edited the manuscript

Co-Author: Matthew W. Fields

Contributions: Principal Investigator, Synthesized ideas, wrote and edited the manuscript

Manuscript Information Page

Anna J. Zelaya, Albert E. Parker, Kathryn L. Bailey, Ping Zhang, Joy Van Nostrand,
Ning Daliang, Dwayne A. Elias, Jizhong Zhou, Terry C. Hazen, Matthew W. Fields

Status of Manuscript:

☒ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☐ Published in a peer-reviewed journal

ABSTRACT

Understanding the factors that determine microbial assembly, composition, and function in subsurface environments are critical to assessing contributions to biogeochemical processes such as carbon cycling and bioremediation. However, these factors are still not fully understood. In this study, surrogate sediment samples were incubated for three months in three wells (FW301, FW303, FW305) within the background site of the Oak Ridge Field Research Center in Oak Ridge, TN. Local surrogate sediment biofilm communities that established were compared to groundwater diversity. Groundwater samples from each well were collected approximately 3 times a week in order to identify members of planktonic communities. Field well geochemistry was also measured. Wells had similar median values for oxidation-reduction potential but not for conductivity, pH, and dissolved oxygen. Additionally, the temporal stability of conductivity, pH, and dissolved oxygen was different for each well. Multiple sediment samples were used per well to determine inter- and intra-well variation for the established bacterial communities. Community analysis of local and source diversity via ss-rRNA paired-end sequencing and distribution-based clustering revealed higher richness, diversity, and variability in source groundwater communities compared to sediment-associated communities. In sediment samples, 20-40% of the sampled populations were abundant at less than 5% within corresponding source groundwater. Inter-well sediment biofilm (across wells) community structure was also distinct from each other but intra-well sediment biofilms were much more similar (n=12 per well). Ordination analysis revealed sediment biofilm communities were distinct from corresponding groundwater communities, with the exception of FW305,

which showed greater dissimilarity of groundwater communities, and similarity between within-well planktonic and sediment communities. Intra-well population networks (SPARCC) with strong associations (correlation cutoff ± 0.8) revealed that overall there were many more correlations within biofilms than with corresponding groundwater populations, which may have been a consequence of temporal fluctuations within groundwater. These results indicate a selection of local community structure that is filtered yet influenced by the available groundwater diversity, geochemistry, and hydrology.

Introduction

The factors and processes that influence organismal assembly into biological communities is a major goal of ecology. Concerns over biodiversity loss due to increased anthropogenic disturbances such as pollution and climate change have motivated ecologists to investigate and understand the mechanisms that control and maintain macro- and micro-biodiversity (Bellard et al, 2012). There has been a general consensus that a combination of stochastic and deterministic processes influence community assembly patterns to different degrees in varying environments (Zhou and Ning, 2017). Deterministic processes such as environmental filtering and species-species interactions have traditionally been considered to play the central role in community assembly (Chesson, 2000). However, stochastic processes, such as random disturbances, ecological drift, and random births and deaths have also been shown to influence community assembly and dynamics (Hubbell, 2001; Chase and Myers, 2011). The relative influences of each process and the spatial and temporal scales at which the different processes impact community structure and

composition are still not fully understood and are areas of intense research and debate (Zhou and Ning, 2017).

The surveyed microbial diversity in shallow, terrestrial subsurface environments has uncovered a rich array of subsurface microbial life that is taxonomically, metabolically, and physiologically diverse (Locey and Lennon, 2016), and based on current genetic and molecular surveys, the subsurface environment is known to be more dynamic environment than previously appreciated (Griebler and Leuders, 2009, Anantharam et al. 2016, Zelaya et al., submitted). However, the biotic and abiotic factors that determine the distribution, abundance, and diversity within subsurface environments, and how these parameters change over space and time in natural environments, is still not yet fully understood. It is usually assumed that soil-associated populations are distinct from and more active than groundwater populations, due to factors such as surface area, nutrient concentration, and biomass retention. However, the true relationships between surrounding diversity in the groundwater and local diversity in the soil are difficult to ascertain based upon logistics of sampling intact soil material in the native state. The major logistics include expensive sampling that only provide a single time and space point that cannot be replicated and the subsequent impact on water flow through the disturbed matrix. When considering the boundary conditions between the bulk-aqueous phase that moves through a porous media, the movement of microbial cells between and along boundaries most likely responds to geochemical perturbations that change temporally and spatially.

Previous studies at the Oak Ridge Field Research Center have shown that pristine groundwater can have a high temporal and spatial variability across short time-scales of

days (Zelaya et al, submitted). Additionally, it has been shown that for a newly cored well, distinct population shifts were observed in a time-sensitive manner that was not observed in older wells. However, these results only described the planktonic fraction of the studied aquifer. In order to obtain a more complete picture of the subsurface diversity (both planktonic and surface-associated communities), surrogate sediment samplers were introduced into the same field wells and incubated concurrently during the 3-month groundwater sampling period. The incubation period served as an opportunity for microbial populations from the surrounding groundwater to colonize surrogate sediment particles and assemble into sediment-associated communities. This study directly compares results from the groundwater study presented in Chapter Two of this dissertation with corresponding surrogate sediment samples. As such, some previously presented data will be revisited in order answer the following questions: 1) how representative were sediment-associated communities to surrounding planktonic fraction after a 3-month incubation period? 2) How (dis)similar were sediment associated communities between three different field wells in close proximity to one another?, 3) What assembly processes (deterministic vs. stochastic) can be inferred from the corresponding SSU rRNA gene taxonomic data.

Materials and Methods

Site Location and Field Wells Chosen

The EOR-FRC background site covers an area of approximately 1.63 km² and is located in the Tennessee West Bear Creek Valley, about 2 km away from any contaminated areas (Appendix A, Supplementary Fig. A.1). The background area is composed of nearly

identical geology to contaminated areas as it lies directly along the same geologic strike. No known contaminants have been disposed at the background area throughout the history of DOE operations. Three sampling wells in the background site (FW301, FW303, FW305) lie in close proximity to each other (~10m apart) and were chosen for this study (see Supplementary Information for details on chosen field wells). Wells FW301 and FW303 are considered historic wells and were installed over 20 years ago. FW305 was a newer well, installed a few weeks prior to the start of this study.

Bio-trap® and Groundwater Sampling

Bio-trap® samplers (Microbial Insights, Inc., Rockford, TN) were used for surrogate sediment incubation (Fig 3.1). The surrogate sediment samplers are made of PVC material with slits cut on either side to allow for water and microbial cells to flow through the chamber. Eighteen total Bio-trap® samplers were lined with either 100µm or 300µm nylon mesh and filled with either 10g of FW301 (numbered 1-12) or FW305 (numbered 13-18) native sediment (that had been stored at -80°C since the time of well coring) and capped with butyl rubber stoppers. Stoppers were secured in place with small, sterilized plastic zip-ties. Samplers 1-12 were randomly grouped into a cluster of six for deployment into either FW301 or FW303. Samplers 13-18 were grouped for deployment into FW305. Teflon thread was used to tie grouped samplers together and for lowering sampler clusters into each well. Tied surrogate sediment sampler clusters were sterilized by autoclaving prior to deployment. The samplers were placed approximately 0.6 m below water table of respective well and incubated for a total of 11 weeks (end of July-mid October). After the incubation period, the samplers were retrieved and transported on ice, at which point each

sample was individually homogenized by pulverization. Samples were stored at -80°C until nucleic acid extraction. Prior to pulverization each bio-trap was cut in half to represent two duplicates (A and B) for each Bio-trap®. Groundwater samples were also collected approximately three times a week during the three-month period for nucleic acid extraction of the planktonic fraction of each well, as well as geochemical characterization (dissolved oxygen, temperature, pH, oxidation-reduction potential, conductivity, sulfate, nitrate, chloride, fluoride). Handling and processing of all groundwater samples has been previously described (Zelaya et al., submitted).

Nucleic Acid Extraction, Sequencing, Quality Filtering, and Taxonomic Identification

For surrogate sediment samples, sediment from each individual sampler was divided into two replicates (A and B for each sampler), resulting 36 total samples for surrogate sediments. However, one replicate was lost during processing (FW-303, sampler 11B), and as such was not included in further analyses. Nucleic acids were obtained using the Modified Miller Method of nucleic acid extraction (Hazen, 2010). The SSU rDNA gene sequences were amplified using a primer pair targeting the V4 region (Forward primer, 515F, 5'-GTGCCAGCMGCCGCGGTAA-3'; reverse primer, 806R, 5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso et al 2010, Lauber et al. 2012). PCR amplification was performed in triplicate and prepared for sequencing according to the MiSeq™ Reagent Kit Preparation Guide using a 500-cycle v2 MiSeq reagent cartridge (Illumina, San Diego, CA, USA).

Raw sequence data was processed using an in-house pipeline (Alm Laboratories, SmileTrain, <https://github.com/almlab/SmileTrain/wiki>). Reads were merged, removed of

primers, quality-score checked, de-replicated, and checked for chimeras. USEARCH was used to merge reads, filter low quality reads, and remove chimeras (Edgar, 2013). Operational Taxonomic Units (OTUs) were clustered using Distribution Based Clustering (Preheim et al., 2013). Taxonomic annotation of individual OTUs were obtained with the RDP 16S Classifier (Wang et al., 2007). Only OTUs with classifications at the Domain level of greater than >80% confidence were used in downstream analyses. The OTUs below this cutoff were removed. OTUs at lower than 80% confidence at the phylum level were named “Unclassified Bacteria” or “Unclassified Archaea”. Additionally, all sequences classified as “Chloroplast” that were not related to Cyanobacteria were removed. Rare OTUs as described by (relative abundance $<10^{-6}$ across all sequences) were removed as described by Haegeman et al. (Haegeman et. al. 2013). Representative sequences were used to build phylogenetic tree by FastTree2 (Price et al 2009, Price et al 2010) after aligned by Clustal Omega (version 1.2.2) (Sievers et al 2011).

Downstream Community Analysis

Diversity profiles based on Hill Numbers was calculated as previously described (Jost 2006). Sorensen and Bray-Curtis pair-wise dissimilarities were determined in R with package VEGAN (Oksanen et. al., 2013). Non-metric multi-dimensional scaling (NMDS) ordinations were constructed using metaMDS in package VEGAN. Correlations for network associations were calculated using sparse correlations for compositional data (SparCC) (Friedman and Alm, 2012) and visualized in Cytoscape (Shannon, et. al, 2003). To remove potential spurious associations due to autocorrelated time-series data of groundwater samples, median values of all filtered groundwater and surrogate sediment

samples per well were used for SparCC analysis. Additionally for network analysis, data was further filtered to be less than 50% sparse in order to remove potential biases of highly sparse datasets (Weiss et al., 2016). For interpretation of association networks, the OTU sediment:groundwater ratio was calculated as the proportion of the median relative abundance of an OTU found across all groundwater and sediment samples of a given well.

Mechanisms Underlying Community Assembly

The relative contributions of community assembly processes were quantified using a framework proposed by Stegen *et al* (Stegen et al 2013, Stegen et al 2015). The influence of selection was estimated based on nearest taxon index between communities (β NTI). A turnover between two communities was considered to be governed by “variable” or “homogeneous selection” when the phylogenetic dissimilarity (β MNTD, beta mean nearest taxon distance) was significantly higher (β NTI>2) or lower (β NTI<-2) than null expectation (Fine and Kembel 2011, Stegen et al 2015, Webb et al 2008). For spatial datasets (*i.e.* surrogate sediments), any turnovers not governed by selection were analyzed using the Raup-Crick metrics (RC) based on Bray-Curtis dissimilarity index to estimate the influence of dispersal (RC>0.95 indicates “dispersal limitation, and RC<-0.95 indicates “homogenizing dispersal”) (Chase et al 2011, Stegen et al 2013). Any process not falling within β NTI or RC ranges were named “undominated/stochastic” (Stegen et al 2015). The relative role of a process was measured as the percentage of turnovers governed by that process between groups (Stegen et al 2013).

Results

Sampled Groundwater Communities Are More Rich and Uneven Than Corresponding Sampled Sediment Communities

Downstream quality filtering of raw paired-end sequence data resulted in a total 990,545 high quality sequence reads (Appendix B, Table B.1) and a total of 3181 unique OTUs. Groundwater samples collectively had 3,167 unique operational taxonomic units (OTUs), while sediment samples collectively had 1,743 unique OTUs (Fig 3.1, Table 3.1).

Diversity profiles based on Hill Numbers (also known as effective species number) is viewed as a measure of true diversity (Jost, 2006) because it shows the relative contributions of evenness and richness to sampled community diversity. The order of q determines how much weight the measure gives to each species' frequency. When q equals zero, the measure is completely insensitive to frequency and the solution is equivalent to species richness. An order of 1 for q results in a measure that weighs each species exactly by the number of occurrences, and the solution is equivalent to the exponential of the Shannon Entropy. When q equals 2, the result is the inverse Simpson concentration, which puts a larger weight on more abundant OTUs. Plotting the resulting Hill Number according to increasing levels of q results in a diversity profile that measures changes in diversity based on the amount of dominance within the community. A perfectly even community would display a flat slope as the value of q increases, while increasing degree of dominance would result in a diversity profile that drops steeply as q increases.

Based on calculated Hill Numbers, overall OTU richness and unevenness for groundwater communities were higher than for sediment sampled communities (Fig. 3.2). FW301 had the highest degree of richness and unevenness out of all groundwater samples. When directly comparing groundwater and sediment communities within the same well, average OTU richness for FW301 and FW303 groundwater were roughly 3 and 4 times higher than for their respective sediment samples, respectively. FW305 groundwater richness was about 1.5 times higher than sediment samples of the same well. Additionally, within-well sediment communities were more even and, with respect to the single time point sampling, highly similar (*e.g.*, high intra-well similarity for sediment associated communities).

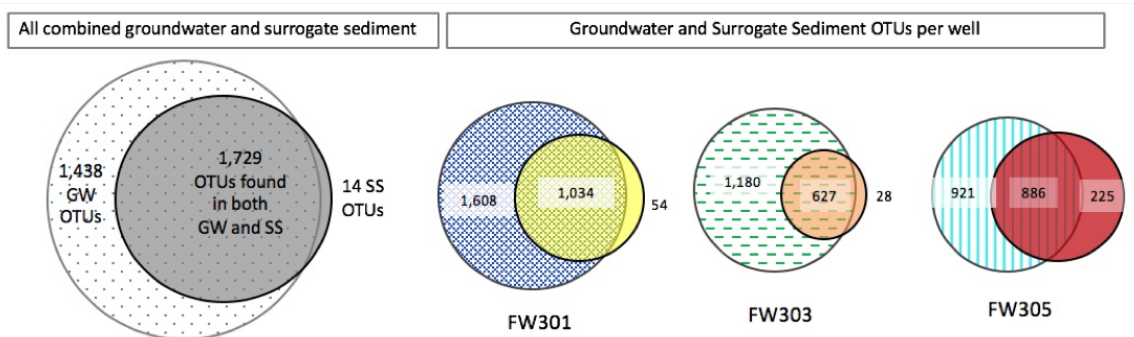


Figure 3.1. Venn Diagrams of total unique OTUs present in sampled groundwater and surrogate. A) The number of unique and overlapping OTUs found in all sampled groundwater and surrogate sediments from all field wells. B) The number of unique OTUs found in sampled groundwater surrogate sediments of FW301, FW303, and FW305.

Table 3.1. Number of OTUs sampled in groundwater and surrogate sediment per well.

	FW301	FW303	FW305
Groundwater OTUs	2642	1926	1807
Surrogate Sediment OTUs	1807	655	1111
Shared in GW and SS	1,034	627	886
Total OTUs	Groundwater 3167	Surrogate Sediment 1743	

Table 3.2. Average sampled groundwater geochemistry for three background field wells over three-month incubation period for surrogate sediments.

	FW301		FW303		FW305	
	Average	SD ±	Average	SD ±	Average	SD ±
Temp (°C)	17.53	0.71	17.85	1.12	17.14	0.68
pH	6.34	1.10	7.29	0.57	6.84	0.64
ORP (mV)	236.82	89.73	239.37	122.17	210.45	69.48
DO (mg/L)	1.64	0.55	0.83	0.40	1.27	0.25
Cond (µS/cm)	172.76	68.72	303.45	26.71	244.75	40.18
Acetate (µM)	1.26	1.28	1.33	1.36	1.33	1.28
Propionate (µM)	0.77	0.66	0.94	1.23	0.78	0.71
Fluoride (µM)	1.47	0.67	4.38	0.89	4.53	1.05
Chloride (µM)	33.86	5.47	30.91	6.77	29.27	4.17
Nitrate (µM)	1.52	1.26	3.62	1.12	2.56	0.91
Sulfate (µM)	63.74	11.30	56.21	14.25	45.74	9.68

Total sampled groundwater diversity included 32 phyla (28 bacterial and 4 archaeal) and 144 families (Appendix B, Supplementary Table B.2). Major populations observed in groundwater were Proteobacteria, Actinobacteria, Acidobacteria, Bacteroidetes, Firmicutes, Planctomycetes, and Verrumicrobia (Appendix B,

Supplementary Fig. B.2), with periodic blooms of Firmicutes in all three wells. In two of the three wells (FW301 and FW305), blooms in Firmicutes were accompanied by a dramatic decrease in total representation of Proteobacteria and an increase in relative abundances of Deinococcus-Thermus. There was one bloom of Fusobacteria in FW303 that did not coincide with a decrease in Proteobacteria representation. Despite these blooms and the dynamic nature of groundwater diversity, sediment samples showed a predominance of Proteobacteria, comprising 95% or more of all sediment samples. Interestingly, FW305 (the youngest well) showed a much larger representation of Bacteroidetes than did sediment samples of the other two wells.

Sediment and Groundwater Communities Are Distinct

Of the 33 phyla classified in RDP (>80% confidence) in sampled groundwater, 8 were completely absent from surrogate sediment samples (Aquificae, Candidate division WPS-2, Cloacimonetes, Hydrogenedentes, Microgenomates, SR1, and Synergistetes) (Appendix B, Supplementary Table B.2). In groundwater, these 8 phyla were observed at very low abundances of the sampled diversity, ranging from 0.002 to 0.05% relative abundances across all samples. At finer taxonomic resolution, only 11 RDP classified families were observed in all sediment samples harvested from all three wells (Fig. 3.3). Shared surrogate sediment families belonged primarily to the Proteobacteria (α , β , γ) with the exception of *Chitinophagaceae* of phylum Bacteroidetes. Cumulatively, these 11 families were found in relative abundances between approximately 30-80% of sampled communities.

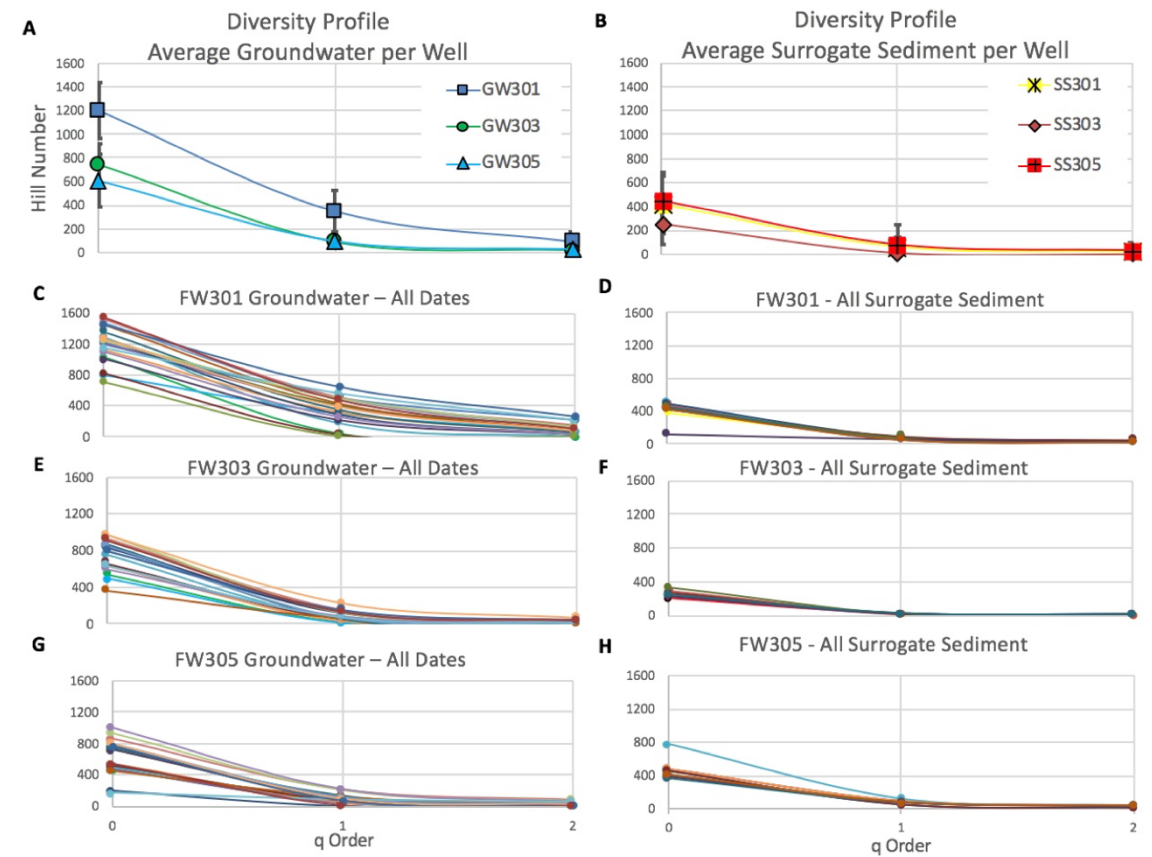


Figure 3.2. Diversity profile of groundwater and surrogate sediment communities based on Hill Numbers. The x-axis represents the order (q) in the Hill Number equation. The y-axis represents the calculated Hill Number of order q . The Hill Number (the equivalent number of equally abundant species) corresponds to species richness when $q=0$. As the value of q increases, the Hill Number is increasingly affected by abundant species. Therefore, linear graphs with flat slopes as q order increases represent perfectly even communities, while lines with decreasing slopes as q increases indicate uneven assemblages. Within-well sediment communities (D,F,H) display similar diversity profiles, indicating high sampling reproducibility, while within-well groundwater communities (C,E,G) vary in their evenness over time. Averages of bio-trap (B) and filtered groundwater (A) samples per well were calculated with error bars indicating standard deviations. Overall, groundwater communities display higher degree of unevenness than corresponding sediment communities, with groundwater samples of well FW301 demonstrating the highest level of unevenness.

While the 11 shared surrogate sediment families comprised the majority of most within-well sediment communities, each well had a distinct profile with respect to relative proportions of these populations. For example, for each bio-trap community within FW301, the cumulative abundances of 5 out of the 11 families represent between 60-70% of the total community (*Comamonadaceae*, *Moraxellaceae*, *Oxalobacteraceae*, *Pseudomonadaceae*, and *Rhodocyclaceae*). In contrast, 2 families cumulatively comprise roughly 30% of community abundances in surrogate sediment samples of FW303, *Pseudomonadaceae* and *Rhodocyclaceae*. Additionally, there is a noticeable decrease of *Comamonadaceae*, *Oxalobacteraceae*, and *Moraxellaceae* populations as compared to those in FW301. A clear and distinct shift of community profiles can be observed in FW305, with the larger representation of *Oxalobacteraceae* populations and a reduction in abundance of *Pseudomonadaceae* as compared to the other two wells. While FW301 and FW303 have a large representation of *Pseudomonadaceae*, populations of this family were observed at roughly 2% in all samples of FW305.

Interestingly, these same 11 families that dominate sampled attached communities were observed in non-dominant proportions in their respective planktonic communities. For example, FW301 bio-trap communities have high representation of *Moraxellaceae*, despite this family not being observed at greater than 5% in all FW301 sampled groundwater communities. This same trend can also be observed in *Rhodocyclaceae* populations within FW303. While comprising between 10-20% of surrogate sediment communities, *Rhodocyclaceae* was only observed in high abundance in groundwater on

the first sampling date (~19% relative abundance). *Rhodocyclaceae* populations decreased to roughly 2-5% relative abundances in all subsequent groundwater samples.

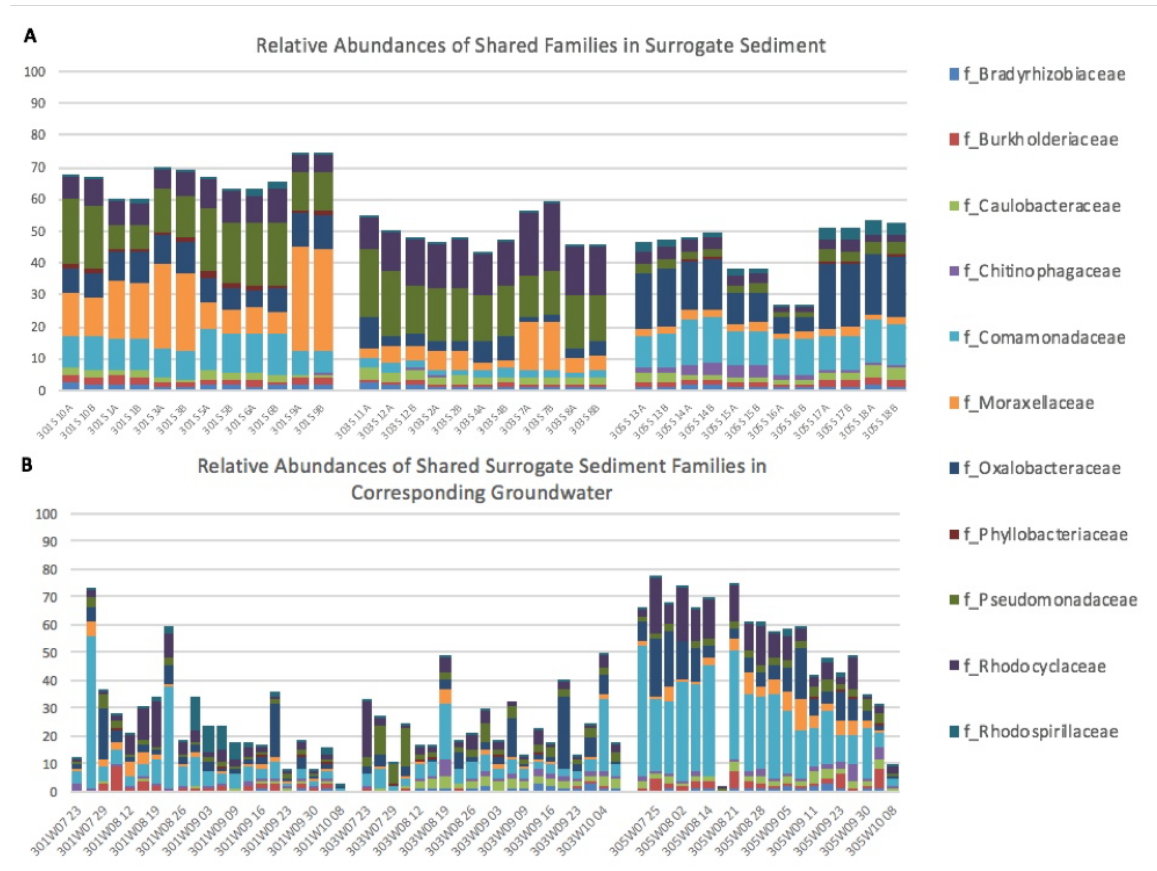


Fig. 3.3 A) Stacked relative abundance bar graphs of the 11 families present in all 36 bio-trap samples, and B) their corresponding relative abundances in sampled groundwater.

Observed shifts in community composition in sampled groundwater over the three-month sampling period indicate that many populations are transiently dominant over time. At the phylum level, periodic blooms of non-Proteobacteria taxa can be observed on different days. For example, Firmicutes blooms can be observed once in FW301 and twice in FW305. FW303 has a smaller bloom of Firmicutes, but has a significant bloom of

Actinobacteria populations on several days. In addition, FW303 groundwater samples contain the highest proportion of Actinobacteria populations out of all three wells over the three-month sampling period. Differences in abundances between local and groundwater communities at finer taxonomic resolutions (OTU) are further described in Figure 3.4 A,B.

Based on non-metric multi-dimensional scaling, most samples cluster uniquely by type and by well, resulting in six unique clusters (Fig. 3.5). Along the primary axis, clusters separate according to type, with distinct clusters that represent groundwater and sediment samples. The secondary axis appears to differentiate samples by well, however the broad spread of FW305 groundwater along both primary and secondary axes results in some overlap between corresponding surrogate sediment samples of that well along with samples from FW303 groundwater and sediment. By contrast, sediment communities from all wells form much tighter clusters, indicating high within-well similarity for sediment communities. Additionally, while sediment clusters tend to separate along the primary axis away from their corresponding within-well groundwater communities, this is not the case for sediment samples of FW305, which groups much more closely to planktonic communities of its groundwater samples.

Higher Taxonomic Resolution Reveals Predominance of Unique OTUs Per Well

While sediment associated communities in all three wells do have populations that are classified within the same taxonomic group (family level, genus level, etc), each well contains its own unique OTU variant of that group (Fig 3.4A). For example, the genus

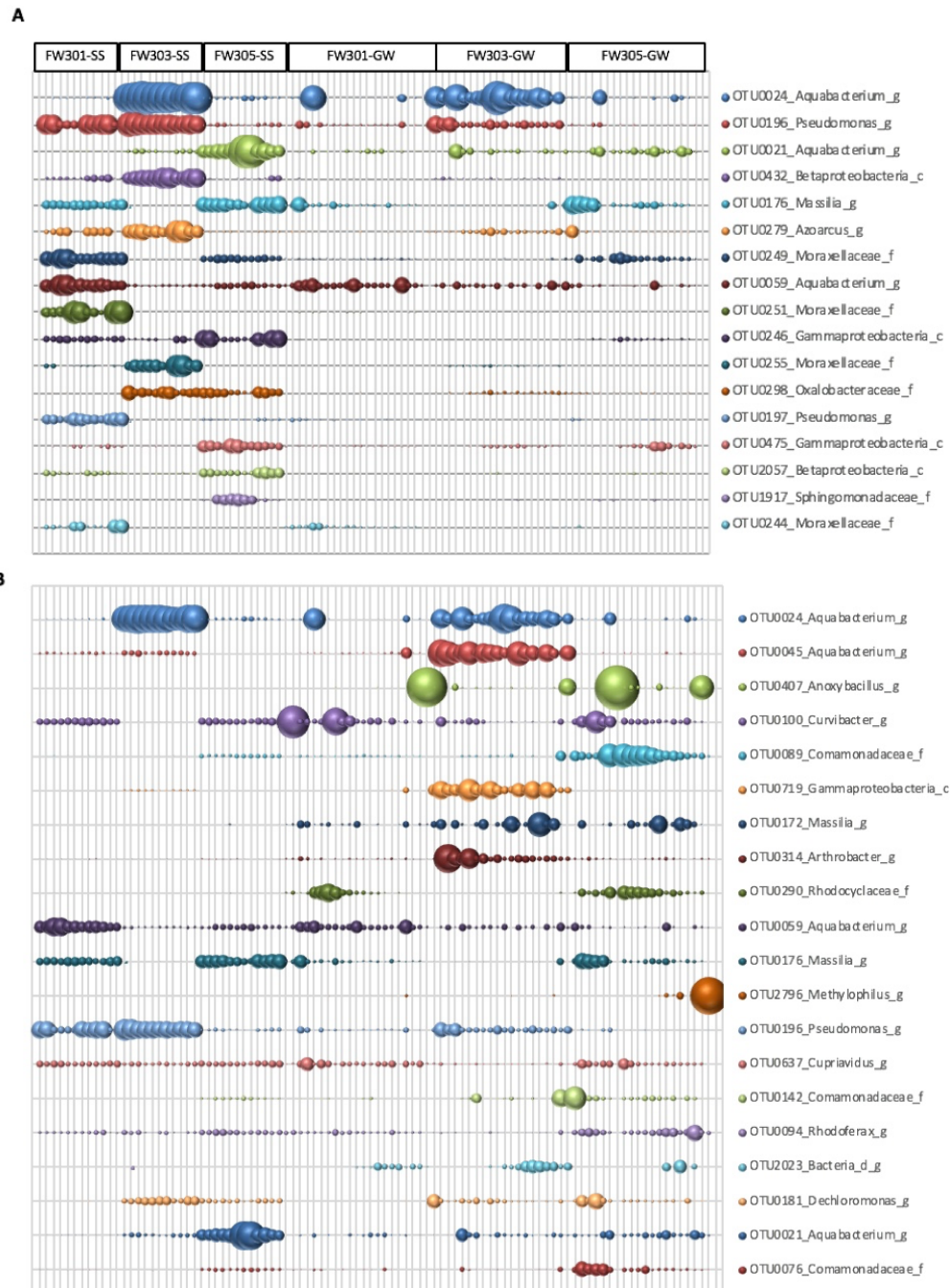


Figure 3.4 A) Most abundant OTUs in surrogate sediments (SS) and their corresponding abundance in filtered groundwater (GW). The timescale for GW samples begins in July and ends in October. B) Most abundant OTUs in groundwater and their corresponding abundances in surrogate sediment. Those that are found abundant in one environment type are not necessarily the most abundant OTUs found in the other. Most

abundant OTUs were defined as those having greater than 5% relative abundance in at least 1 sample per well. Bubbles size represents relative abundance within sample and were scaled to 25% of original bubble size to reduce overlaps between bubbles.

Aquabacterium is observed in very high abundances in all three wells, however, each well has its own unique OTU variant of *Aquabacterium*; OTU0059 in FW301, OTU0024 in FW303, and OTU0021 in FW305 (Fig. 3.4A). Several other examples of this can be observed for an unclassified member of the family *Moraxellaceae*, β -proteobacteria, and γ -proteobacteria. Unique variants within groundwater communities were also observed (Fig. 3.4B). However, due to the transient nature of groundwater populations, a core-set of populations within each well were more difficult to determine.

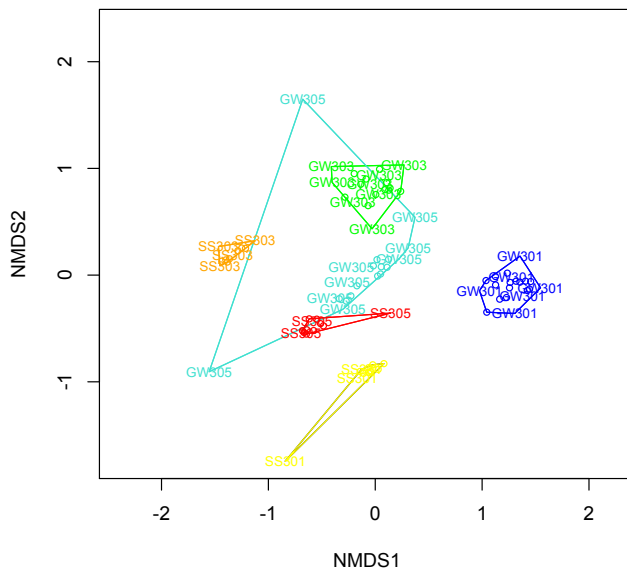


Figure 3.5. Representation of pairwise Bray-Curtis distances between filtered groundwater (GW) and surrogate sediment (SS) samples using two dimensional NMDS ordination. Data converged after 20 random starts with stress value of 0.167. Samples from each well and type are color coded as follows: groundwater from FW301 (GW301, blue), FW303 (GW303, green), FW305 (GW305, turquoise) and surrogate sediment harvested from FW301 (SS301, yellow), FW303 (SS303, orange), and FW305 (SS305, red).

Co-occurrence Networks

Intra-well population networks (SPARCC) with strong associations (correlation cutoff ± 0.8) revealed that overall, each well had distinct associations between OTUs that

were observed more exclusively in sediment or groundwater and vice versa [based on sediment:groundwater (ss:gw) ratio] (Fig. 3.6). Additionally, many more correlations were observed in sediment-associated populations (*i.e.*, biofilm) than corresponding groundwater populations overall. FW301 had the greatest number of total associations (386) with all but one being between OTUs of within surrogate sediment samples. The one negative ss:gw association was represented by sediment-associated β -proteobacteria (unclassified at lower taxonomy) and groundwater-associated α -Proteobacteria (classified to order Rhizobiales). Meanwhile, FW303 had the least total associations (194) with 8% being negative, many of which were between OTUs that appeared to have preferential phases (preferred surrogate sediment over groundwater or vice versa) based on ss:gw ratios of median relative abundances. FW305 had the greatest number of negative associations (16%) between OTUs. Additionally, FW305 had negative associations between OTUs that both preferred surrogate sediment over groundwater and vice-versa based on ss:gw ratios of median relative abundances.

Recently, methods for estimating relative contributions of neutral and niche-based processes towards community assembly have been introduced (Stegen et. al, 2015). Using β NTI to calculate relative contributions in the present study, homogenous selection seems to be a major contributing process in all samples compared (Fig. 3.7). Within samples of surrogate sediment, it may be the only process involved, suggesting strong biotic or abiotic pressures select for closely related taxa from one surrogate sediment sample to the next. In filtered planktonic assemblages of sampled groundwater, each well experienced different contributions of both neutral and niche-based processes. FW305 was mostly explained by

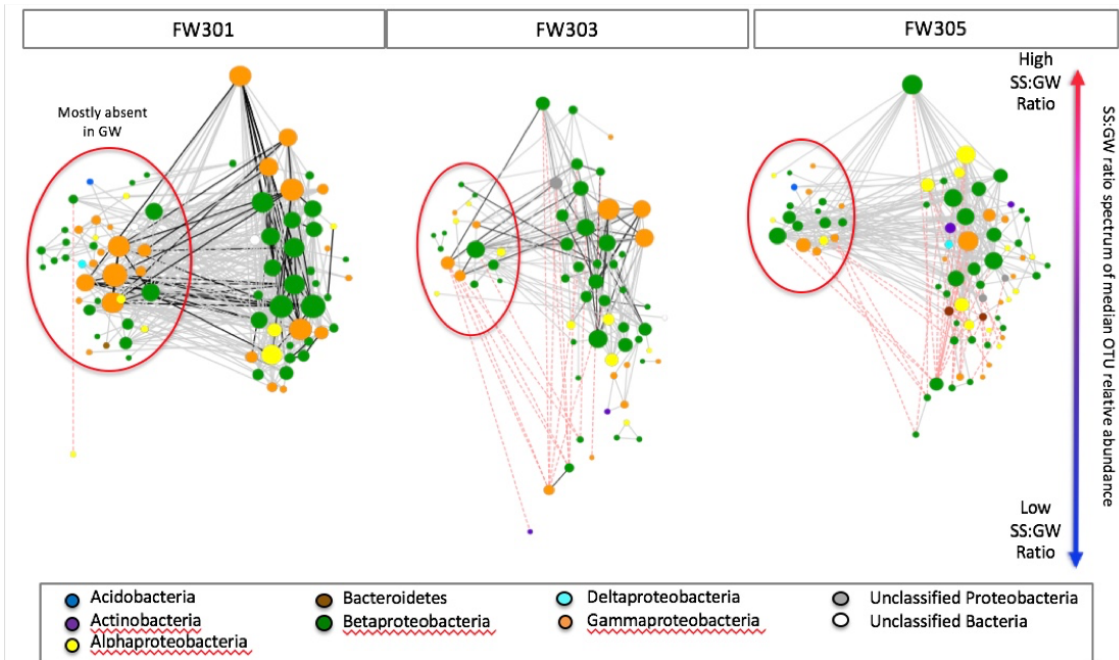


Figure 3.6. Community networks of co-occurring OTUs based on correlation analysis of within-well groundwater and sediment samples with SparCC. Positive associations are indicated by a solid grey line. The strength of the association is indicated by the intensity of the line (black lines have correlation of ≥ 0.9 , gray lines have correlation of ≥ 0.7). Negative associations are indicated by a dashed red line (≤ -0.7). All associations are significant ($P\text{-value} = 0.05$). Nodes are colored according to the RDP taxonomic classification ($>80\%$ confidence level) at the Phylum level except for Proteobacteria which are further denoted by Class. The size of each node is proportional to the number of connections with other nodes (degree). Nodes found towards the center indicate similar median relative abundances in both groundwater and surrogate sediment samples. Nodes shifted to the bottom have higher median relative abundances in groundwater than in corresponding sediment. Nodes shifted to the top have higher median relative abundances in surrogate sediment than in corresponding groundwater. Graphical distances are not to scale. Some ratios could not be calculated because median relative abundance values in opposite type equaled 0 (described as “Mostly absent”).

homogenous selection, with roughly 10% undominated (stochastic) process contributions.

Homogenous selection in FW303 comprised about 75% of all explanatory processes, homogenous dispersal and undominated (stochastic) processes cumulatively accounting for about 12.5%. In FW-301 groundwater, homogenous selection comprised nearly 65% of all processes, followed by undominated (stochastic) processes accounting for roughly

35% of all processes. When comparing communities across groundwater and surrogate sediment within each well, FW303 seemed to be the most affected by undominated/stochastic processes, with nearly half of processes being undominated/stochastic.

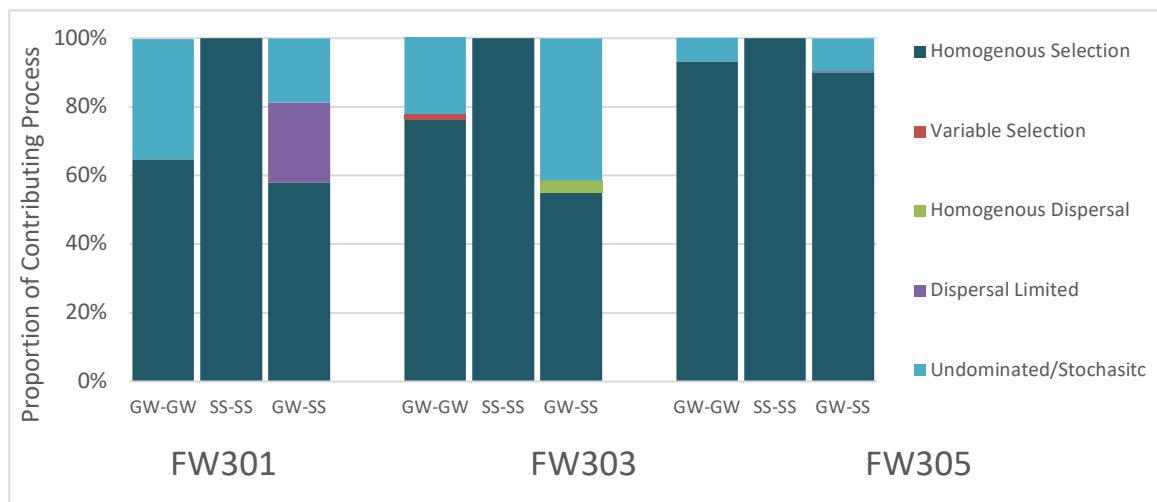


Figure 3.7. Relative influence of community assembly processes on spatiotemporal turnover of within-well bacterial assemblages. Processes were calculated using RC_{Bray} and bNTI as described in materials and methods.

Discussion

Understanding community assembly dynamics is an important goal of microbial ecology. In this study, sterile surrogate sediment samples were incubated and colonized by groundwater microbial cells within an uncontaminated, oligotrophic aquifer in Tennessee. Community composition between local (surrogate sediment) and groundwater communities were compared. Key differences between sediment associated and groundwater populations were observed. Overall, both groundwater and sediment samples

were dominated by α -, β -, and γ - Proteobacteria. However, groundwater displayed greater richness and diversity that fluctuated over time uniquely in each well.

In terms of initial community assembly, the influence of early colonizers due to founder effects is believed to play an important role. Experimental approaches using microcosms to examine the roles of stochastic versus deterministic processes in community assembly have shown groundwater diversity may have a greater effect on bacterial community composition than environmental conditions in early assembly (Langendhede 2006, Langendhede and Szekely, 2011). Additionally, initial colonization of microcosms may occur in a stochastic manner, followed by deterministic processes that lead to filtering and selection which help to shape community structure in the long term (Zhou et al, 2012, Lee et al. 2013).

For this study, it is assumed that the founding members of local surrogate sediment communities were introduced via the populations within the groundwater of each respective well. However, as this study did not compare surrogate sediment associated community changes over time, it is not possible to determine temporal compositional shifts of attached communities. At the end of the 3-month incubation period, different populations (OTUs) of taxa similar to those observed in groundwater were associated to surrogate sediment. The specific surrogate-sediment associated OTUs were often observed as rare populations within the groundwater diversity, and their abundances within the groundwater fraction often fluctuated over time. Additionally, the most abundant populations in groundwater tended to fluctuate over time and were often observed as rare or absent members in corresponding surrogate sediment at the final time point.

The high degree of temporal fluctuation of groundwater communities over short periods (days) was an initially surprising result (Zelaya et al., submitted). In subsurface aquifer environments, immigration and dispersal are believed to be driven in part by the hydrogeological mixing and flow (Smith et al., 2018). In various subsurface systems, hydrogeological circulation has been shown to have a strong effect on microbial communities by affecting groundwater residence time, rates of bacterial transfer, hydrochemistry, and nutrient supply (Ben Maamar, 2015, Bougon et al., 2010, Bougon et al., 2012, Zhou et al., 2012). It is possible that fluctuating abundances within groundwater were due to transient immigration/emigration of certain populations via hydrological flow, or as a consequence of surrogate sediment surface attachment/detachment over time, perhaps due to responses of fluctuating hydrogeochemistry and nutrient availability. Hydrologic parameters required to estimate *in-situ* residence times that would provide insight on the transient nature of groundwater populations can difficult to measure (Watson et al., 2004). Based on crude estimates calculated with existing hydrological measurements (Watson et al., 2004), residence times within the EOR-FRC can vary between 3.9 to 390 hours (Appendix A, Supplementary Table A2). At short residence times, transient populations may have been introduced via vertical or lateral transfer of groundwater. Temporal fluctuations of certain geochemical parameters (*e.g.* DO, ORP, chloride) were significantly associated with changes in planktonic richness within the two older wells (Zelaya et al., submitted). While temporal geochemical fluctuations could not be directly related to the one-time end-point sampling of surrogate sediment, the attempt to associate possible connection with geochemistry and surrogate sediment communities based on the

geochemistry values of each well on the day of sampler harvesting was made via a constrained ordination analysis (Supplementary Figure B. , Supplementary Table B.3). Overall, highly associated populations belonged to α -, β -, and γ -Proteobacteria in all three wells. It is possible that fluctuations in certain geochemical parameters, for example anions and Conductivity, may have influenced the attachment and detachment of surface associated populations over time. Future work is currently under way to more intentionally study attachment/detachment dynamics of planktonic and surface-associated communities at the solid/liquid interface within in controlled laboratory systems (*i.e.* modified packed-bed reactors) that can better control hydrogeochemical parameters.

In ecology, filter effects describe a type of community assembly process by which biotic or abiotic factors prevent the establishment or persistence of a species in a particular location (Keddy, 1992)*. Filter effects have been widely used to describe community turnover between various environments (Thakur and Write, 2017, references there-in), and is believed to play important roles in community assembly (Loreau et. al, 2001, Freedman et. al, 2015). Recently, the concept of filter effects as a major explanation for compositional differences has been criticized (Kraft, 2015), largely because it is insufficient to differentiate between the biotic and abiotic contributions towards possible selection. Of note, competitive exclusion as described by coexistence theories has been emphasized to highlight the insufficiency of the environmental filter hypothesis (Kraft et. al., 2015).

Recently, methods for estimating relative contributions of neutral and niche-based processes towards community assembly have been introduced (Stegen et al., 2015). Using β NTI to calculate relative contributions in this study, homogenous selection seems to be a

major contributing process in all samples compared. Interestingly, comparisons between groundwater and sediment communities showed differing contributions of processes towards community assembly, indicating that each well environment allows for differing processes that control community assembly. In the youngest well, FW305, which represents an environment having recently undergone a disturbance event (well-construction), homogenous selection was the primary process within and between groundwater and surrogate sediment, with minor contributions of undominated (stochastic) processes. These findings corroborate studies that showed stochastic processes and priority effects are important in initial assembly of waters and sediments (Crump et al 2012, Hao 2015). However, over time (in this case over the three-month period) populations within communities may undergo selection via biotic or abiotic pressures (Zhou et al, 2012, Lee et al. 2013). This explanation is also corroborated by the dramatic shifts in the community composition observed in groundwater communities over time in FW305 that were not observed in the other two wells. In contrast, while the dominant members of FW305 groundwater communities shifted over time, in FW-301 it was the variability of rarer and less abundant populations that accounted for the majority of compositional shifts. This may explain why out of all three wells, FW301 was the most affected by undominated, or this well. It is possible that these processes may have included random births/deaths (*e.g.* cell growth or lysing events), or immigration/emigration of transient populations. FW303

*Note: Within the context of community assembly dynamics, the phrase “filter” is a term used by ecologists to describe environmental constraints as “filters”, and should not be confused with the polyethersulfone (PES) filters used for biomass collection in Materials and Methods.

stochastic, processes (22%), which agrees with the high temporal variability observed in was the only well where homogenous dispersal was observed (11%), which may be associated with overall geochemical and temporal stability (temporally homogenous environment) observed in that well compared to the other two wells.

Several studies of both macro- and micro-organisms have demonstrated that a combination of neutral and niche-based processes most likely drive community assembly in nature (Zhou and Ning, 2017). For microorganisms, it has been hypothesized that neutral based processes may have a larger role in community assembly due to microbial properties that are distinct and unique from macro-organisms (*e.g.* small sizes, dispersal capabilities, growth rates) (Farjalla et al., 2012). For this study it was observed that subsurface microbial populations of an uncontaminated aquifer may be influenced uniquely by both neutral- and niche-based processes across the groundwater/sediment boundary, and that surrounding groundwater community composition as well as environmental factors help to shape community structure over space and time between source and local diversity for a shallow, porous media with groundwater flow.

These results presented in this chapter provide insight into the possible community dynamics between subsurface bacteria observed within planktonic and solid-associated states. Across the surrounding groundwater and surrogate sediment (local) environments, the results presented here suggest selection of local community structure that is filtered yet influenced by the available groundwater diversity, geochemistry, and hydrology.

CHAPTER FOUR

COMMUNITY STRUCTURE ACROSS PARTICLE SIZE IN-VITRO AND IN-SITU

Contribution of Authors and Co-Authors

Manuscript in Chapter Two

Author: Anna J. Zelaya

Contributions: Performed literature review, synthesized ideas, performed DNA extractions and paired-end sequencing, performed statistical and bioinformatical analyses, wrote and edited the manuscript

Co-Author: Sara Altenburg

Contributions: Synthesized ideas, maintained and sampled laboratory bioreactors, data collection

Co-Author: Kathryn L. Bailey

Contributions: Performed field samplings, geochemical analysis and data collection, reviewed and edited the manuscript

Co-Author: Heidi J. Smith: Performed EEMS analysis, wrote methods

Co-Author: Dwayne A. Elias

Contributions: Supervisor and Principal Investigator to Kathryn L. Bailey

Co-Author: Matthew W. Fields

Contributions: Principal Investigator, Synthesized ideas, wrote and edited the manuscript

Manuscript Information Page

Anna J. Zelaya, Sara Altenburg, Kathryn L. Bailey, Heidi J. Smith, Dwayne A. Elias,
Matthew W. Fields

Status of Manuscript:

☒ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☐ Published in a peer-reviewed journal

ABSTRACT

Global biogeochemical processes are mediated by microorganisms that are associated with solid-phase particulates within subsurface environments. Subsurface sediments can be heterogeneous with respect to mineralogy, organic matter content, and particle size. Currently, the impact of physical surface scale on subsurface microbial diversity and distribution is poorly understood. In this study, unfiltered groundwater from a non-contaminated area of the Oak Ridge Reservation (ORR) was fed into modified CDC biofilm reactors to assess the microbial community structure on solid particles. Standard coupon holders of an anaerobic CDC biofilm reactor were modified to contain glass beads of various sizes (30 μm , 425 μm and 3,000 μm) with continuous access to nutrients from in-fed groundwater. In order to relate laboratory findings to *in-situ* dynamics, surrogate samplers were filled with glass beads of either (30 μm , 425 μm and 3,000 μm) and incubated down field wells within a non-contaminated area of the ORR. Microbial diversity observed on *in-situ* glass beads was also directly compared to structure observed on native surrogate sediments incubated in the same field wells in a previous study. The results indicate that there may be particle-size effects on community assembly dynamics, but that these effects may be outweighed by other environmental factors unique to each respective well. Future work will build on the results presented here in order to obtain a greater understanding of the interplay between particle size, hydrodynamics, and microbial community assembly and distribution in subsurface environments.

Introduction

The shallow terrestrial subsurface is a complex and heterogenous environment that harbors an important reservoir of microbial biodiversity. While microorganisms have been linked to key biogeochemical processes within subsurface environments (Fredrickson and Fletcher, 2001), it is still unclear how microbial communities are driven by biotic and abiotic factors, and at which spatial resolution those factors work. The ongoing hypothesis relating increased biodiversity to increased ecosystem functioning continues to lack a clear understanding of the spatial scales at which relevant processes occur.

Within subsurface environments, microorganisms reside in a complex matrix of soils, sediments, and particles of various sizes and geochemical properties. The microenvironments that exist at scales relevant to microbes (*e.g.* μm) have been shown to vary significantly in terms of type and content of available organic matter, oxygen concentration, water availability and flow (Nimmo, 2004; Rabbi et al., 2016). Driven by aspects of soil and sediment architecture (*e.g.* porosity, pore-size distribution, pore connectivity), microbes within subsurface environments can experience drastically different microenvironments that vary with respect to the retention and movement of fluids such as water and air, the transport and reaction of chemicals, and biotic stresses such as competition with, and predation by, other macro- and micro-organisms (Nimmo, 2004)

Traditionally, microbial surveys have relied on bulk measurements (*e.g.* grams to kg of soil) in order to infer connections between bacterial diversity and ecosystem functioning (*e.g.* denitrification, carbon sequestration) on large field scales [*e.g.* landscape (Lauber et al., 2012); region (Staley and Sadowsky, 2016); continent (Lauber et al., 2009)].

However, microbial interactions and biogeochemical processes that occur within micro-niches and micro-habitats can be obscured by the homogenization of bulk measurements (Ruamps et al., 2011). The question remains whether micro-scale community structure occurs in a predictable manner, and if so, how that might affect larger scale diversity and function. Studies that have looked at microscale diversity have shown key differences between populations residing on the interior and exterior of micro-aggregates. Studies examining community diversity at the particle-scale have been performed on agricultural soils (Jiang et al., 2018), as well as sediments of freshwater lake systems (Jackson and Weeks, 2008) and marine environments (Mestre et al., 2018). Investigations at these finer resolutions may help uncover patterns in microbial community structure and composition that have consequences at larger scales.

In order to investigate the impact of physical surface scale on microbial interactions, unfiltered groundwater from a uncontaminated area of the Oak Ridge Reservation (ORR) was fed into modified biofilm reactors to assess the microbial community that grew on various particle sizes. Standard coupon holders of an anaerobic CDC biofilm reactor were modified to contain glass beads, with each reactor containing beads representing one of three approximate particle sizes (30 μm , 425 μm and 3,000 μm). Reactors were maintained with continuous access to nutrients from in-fed groundwater. In order to relate laboratory findings to *in-situ* dynamics, surrogate samplers were filled with glass beads of either (30 μm , 425 μm and 3,000 μm) and incubated down field wells within a non-contaminated area of the ORR.

Materials and Methods

Environmental Background Site

The ENIGMA-Oak Ridge Field Research Center (EOR-FRC) contains an uncontaminated background area that has been previously described (Chapter Two, pg. 20). The same three field wells that were chosen in the previous spatiotemporal study (FW301, FW303, FW305) were selected for the present study (see Appendix A, Fig. A.1. for details on chosen field wells). Additionally, for laboratory reactor studies, unfiltered groundwater from well FW301 was collected in a sterile 20L carboy and shipped overnight under refrigeration prior to the start of laboratory reactor studies.

Bioreactor Set Up and Maintenance

Four custom designed anaerobic bioreactors (BioSurface Technologies, Bozeman, MT) were used for the present study. Each bioreactor was designed to hold up to eight removable coupons that were filled with borosilicate beads of desired diameter size. A total of three reactors containing either 1) 30-50 μ m, 2) 425-600 μ m, or 3) 3000 μ m sized borosilicate beads and an additional fourth reactor without any coupons or beads (no-bead control) were set up to run in parallel and be fed by the same groundwater (FW301). To start the study, each bioreactor was filled with 500 mL of FW301 groundwater and allowed to acclimate for 24 hours in batch. After 24 hours, continuous flow was started and groundwater was fed into each reactor using gas impermeable tubing (Masterflex® Norprene) at a flow rate of 0.13 mL/min. Reactors were under continuous stirring and continuous flow to maintain constant mixing of the bulk aqueous fluid for the course of the

study (six months). Additionally, each reactor contained ~500mL of headspace that was continuously flushed with low flow rate of N₂ gas. While dissolved oxygen content was not directly measured during the incubation period, it was assumed that microaerophilic conditions were reached based on previous studies using similar reactor set up. Purely anoxic conditions were not desired as background groundwater has been shown to fluctuate between anoxic and microaerophilic conditions *in-situ*. All reactor components, including beads and tubing, were sterilized by autoclaving prior to the start of the study. A photo of reactor design and set-up can be found in supplementary materials (Appendix C, Supplementary Figure C.3).

Emission Excitation Matrices (EEMs)

Samples for OM (organic matter) spectral characterization were collected from aqueous effluent of each reactor at the end of the six-month incubation period and filtered through 0.2µm low-carbon binding filters into septa sealable quartz cuvettes. Following filtration, samples were immediately analyzed at room temperature. Absorbance measurements greater than 0.3 at 254 nm were diluted with Milli-Q water to prevent inner-filter effects during collection of emission excitation matrices (EEMs) (Miller et. al 2010). EEMs were collected with a Fluoromax-4 Spectrofluorometer, equipped with a Xenon lamp light source and a 1 cm pathlength quartz cuvette. Excitation (Ex) wavelengths were scanned from 240-450 nm and emission (Em) wavelengths were recorded between 300-550 nm in 2 nm increments, with a 5 nm slit width and 0.25 s data integration time. Post processing was completed in MATLAB to generate EEMs corrected for inner filter effects, Raman scattering, and blank water subtraction (McKnight, 2001, Lawaetz, 2009).

Following EEMs analysis samples were analyzed for UV-absorbance (190 nm to 1100 nm) using 190-1100 nm using a Genesys 10 Series (Thermo-Scientific) Spectrophotometer with a 1 cm path length cuvette and Mill-Q water as a blank. EEMs post-processing was performed in MATLAB to correct for inner-filter effects and Raman scattering. A Milli-Q water blank was run each day and subsequently subtracted during post-processing.

Cell Counts From Bulk Fluid of Laboratory Bioreactors

Aliquots of internal reactor bulk liquid were sampled via a top sampling port of the CDC reactor lid. Gas remained continuously flushing the system to limit the potential of atmospheric air introduction into reactors. A total of 3mL was removed from each reactor and carboy for cell counts using Acridine Orange. Triplicate 1mL aliquots were fixed with 37% formaldehyde, followed by the addition 100 μ L of 1% aqueous solution of Acridine Orange stain. Samples were allowed to incubate at room temperature for 15 min. Each mL was then filtered through 0.2 μ m black polycarbonate filter via vacuum filtration. Filters were then mounted onto glass slides and counted using a Nikon epifluorescent microscope equipped with an 100X oil immersion objective. Bacteria were counted until 300 total cells were enumerated or until 30 total fields were counted. Duplicate filters were made for each sample.

Bio-trap® and Groundwater Sampling

Bio-trap® samplers (Microbial Insights, Inc., Rockford, TN) were used for *in-situ* glass bead incubations and have been described in detail previously (Chapter Three, pg. 59). For this study, six Bio-trap® samplers were each lined with 300 μ m nylon mesh and

filled with either 30-50 μm , 425-600 μm , or 3000 μm sized borosilicate beads. For the smallest bead size (30-50 μm), a 30 μm mesh was used to line the sampler to ensure beads remained within the reactor. Each sampler was then capped with butyl rubber stoppers. Stoppers were secured in place with small, sterilized plastic zip-ties. Then two samplers with each bead size were randomly selected and grouped into a cluster of six for deployment into either FW301, FW303, FW305. Teflon thread was used to tie grouped samplers together and for lowering sampler clusters into each well. Tied sampler clusters were sterilized by autoclaving prior to deployment. The samplers were submerged within the groundwater of their respective well and incubated for a total of 12 weeks (April-July). After the incubation period, the samplers were retrieved and frozen at -80°C until nucleic acid extraction. Filtered groundwater samples were also collected approximately once a week during the three-month period for nucleic acid extraction of the planktonic fraction of each well. Additionally, temperature, pH, oxidation-reduction potential, dissolved oxygen, and conductivity were measured as previously described (Zelaya et al., submitted).

Individual and Bulk Bead Collection for Nucleic Acid Extraction

In order to perform community analysis across particle size, both bulk and individual bead fractions were analyzed. Individual beads from both bioreactors and *in-situ* samplers were individually sorted to obtain nucleic acids for a per-bead analysis of community structure. Individual beads of intermediate (425-600 μm) and large (3000 μm) sizes were carefully removed from coupons and BioTrap® samplers using sterile tweezers and transferred to a sterile PCR tube. For the smallest bead size (30-50 μm), beads could not be individually sorted. As such, a subset of beads of this size were carefully drawn

using a sterile 26-gauge needle, and transferred to a small PCR tube. The total number of 30-50 μm beads taken from each draw were counted inside their respective PCR tubes using a stereoscope. Typically <200 beads per draw were sampled using this method (Appendix C, Supplementary Table C.1). Finally, 30 μL of TE buffer was added to each PCR tube, after which tubes were frozen in -20°C until nucleic acid extraction. For bulk bead analysis, approximately 1g of beads from *in-situ* samplers were aseptically transferred to microcentrifuge tubes and frozen at -20°C until nucleic acid extraction.

Nucleic Acid Extraction and 16S rRNA Amplification for Individual and Bulk Beads

For individual beads, a boil method was used for DNA extraction. PCR tubes containing TE buffer and single (425 μm , 3000 μm) or <100 (30-50 μm) beads were boiled at 95°C for 15 minutes. A total of 8 μL was then drawn and transferred to 25 μL PCR reactions containing 1 μL of each 10mM Illumina-barcoded primers (Forward and Reverse), 12.5 μL of HS Premium Bullseye Taq Polymerase (MidSci™, St. Louis, MO), and 2.5 μL nuclease free molecular grade water. The primer design consisted of Illumina adaptor sequences followed by universal forward 8F 5'-tcgtcggcagcgtcagatgtgtataagagacagAGAGTTTGATCCTGGCTCAG-3' or 529R 5'-gtctcgtgggctcggagatgtgtataagagacagACCGCGGCKGCTGGC-3' primers. PCR was performed in an Eppendorf Mastercycler Pro S. The amplification protocol consisted of an initial denaturation at 94°C for 15 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 60 s, extension at 72°C for 60 s, and a final extension step at 72°C for 7 min. Negative PCR controls without DNA template were amplified to monitor contamination. There was no evidence of amplification from any of the tested procedural

blanks and negative sorts and these samples were therefore not included in successive analyses.

For both bulk beads and groundwater filters, the method chosen was used previously for surrogate sediment and groundwater community analysis (Ch. 2, pg. 21 and Ch. 3, pg. 60 of this dissertation). Briefly, nucleic acids were obtained using the Modified Miller Method of nucleic acid extraction (Hazen, 2010) and amplified using the primer pair targeting the V4 hypervariable region of the 16S rRNA gene (Caporaso et al 2010, Lauber et al. 2012). PCR amplification was performed in triplicate and prepared for sequencing according to the MiSeq™ Reagent Kit Preparation Guide using a 500-cycle v2 MiSeq reagent cartridge (Illumina, San Diego, CA, USA). The rationale behind performing similar extraction method as had been done in the prior field study was to facilitate direct comparisons between how previously characterized community structure from surrogate sediment within the same field wells might be influenced by effects of particle size. At the time of this writing, bulk beads and groundwater filters were in the process of being prepared for sequencing, and as such, results for this portion of the study will not be presented or discussed as they are not yet available.

Sequencing and Sequence Analysis

SSU rRNA gene amplicons were sequenced using the Illumina MiSeq Platform (San Diego, CA, USA). The manufacturer's Metagenomic 16S Sequencing protocol for 300-read paired end sequencing was followed. After endpoint PCR amplification, amplicons (containing MiSeq adapter sequences) were purified (AmpureXP beads), indexed, bead purified, quantified with PicoGreen (Quant-IT, Invitrogen, Carlsbad, CA,

USA), and concentrations for each indexed sample was normalized. Following the normalization of DNA concentrations samples were all pooled together, mixed with the PhiX control (10% of sample DNA concentration), and sequenced.

Sequence analysis of the forward and reverse sequences obtained from individual beads were joined with the Quantitative Insights Into Microbial Ecology (QIIME) toolkit version 1.9.0 (Caporaso *et al.* 2010). Sequences quality refinement was performed by the removal of sequences containing ambiguous bases, homopolymers >8 bases, and reads with an average quality score below 30 over a 50 bp window. Operational taxonomic units (OTUs) were chosen using the open-reference method within QIIME and sequences were clustered to 97% similarity. Chimeric sequences were removed using UCHIME (Edgar *et al.* 2011) with the SILVA Gold database (Pruesse *et al.* 2007). A second chimera check was performed using sequences from the present study as a database. Taxonomy was assigned using NCBI database. OTUs classified as “Chloroplast”, containing less than 2 ($n=2$) across all samples, and cumulative sequence counts across all samples representing 10^{-7} or fewer of all total sequences were removed. For bulk beads and groundwater, raw sequences were processed using the same method as described previously (Ch. 2, pg. 22 and Ch. 3, pg. 61 of this dissertation).

Diversity estimates based on Hill Numbers were calculated (Jost, 2006). Sorensen and Bray-Curtis pair-wise dissimilarities were determined in R with package VEGAN (Oksanen *et. al.*, 2013). Non-metric multi-dimensional scaling (NMDS) ordinations based on Bray-Curtis dissimilarities were constructed using metaMDS in package VEGAN. OTUs were arbitrarily classified as Abundant (>1%), Intermediate (between 0.01% and

1%) and rare ($<0.01\%$). One-way ANOVA's, Tukey's Honestly Significant Difference (TukeyHSD) tests, and linear regression models were performed in R statistical base package to discern relationships between bead size and community attributes.

Results

Richness and Diversity of Bead Sizes Differed in an Environment Dependent Manner

A total of 3,395,646 high quality sequence reads representing 8,589 OTUs were observed across all particle sizes and environment types (reactors and field wells). Beads across all samples ranged in total richness from 307-3029 OTUs per bead size (Fig. 4.1). Richness and diversity based on Hill Numbers showed that overall richness and diversity did not differ significantly based on bead size alone (Fig. 4.1). Instead, environment was the biggest predictor of differences between different bead samples (Fig. 4.3), and beads sampled from the same environment were more similar to each other than beads of the same size from different environments (Fig. 4.2).

However, when examining differences between bead size within the same environment, differences between richness and diversity can be observed (Fig. 4.3). In laboratory reactors, richness observed on the largest size beads were significantly higher than either of the other two bead sizes (Fig. 4.4A). Diversity, as based on the exponential of Shannon Entropy (Hill Number, $q=1$), across differently sized beads within bioreactors and within FW301, but not for FW303, were significantly different from each other (Fig. 4.4B). Interestingly, for environments where diversity differed significantly, the size of

Table 4.1. Measured Geochemistry over time for field wells sampled in this study.

Well ID	Day	Temp (°C)	pH	ORP (mV)	DO (mg/L)	Cond (µS/cm)
FW301	4/30/14	15.16	5.28	347	3.4	41.22
FW301	5/5/14	17.23	5.37	175	2.04	54.13
FW301	5/13/14	17.2	7.11	124	1.87	270.7
FW301	5/19/14	14.89	6.29	301	1.41	247.8
FW301	5/28/14	17.13	6.75	125	2.14	289.7
FW301	6/2/14	15.37	6.75	113	2.1	291.3
FW301	6/11/14	16.44	6	345	2.06	312.1
FW301	6/18/14	19.19	5.82	124	1.53	164
FW301	6/24/14	17.63	6.74	112	1.76	147.4
FW301	7/2/14	15.93	6.71	120	1.97	158.3
FW301	7/9/14	17.72	6.71	92	1.48	196.3
FW301	7/16/14	15.87	5.83	239	1.27	235.9
FW301	7/23/14	17.85	6.14	242	2.7	236
FW303	4/30/14	15.36	6.78	262	1.00	255.9
FW303	5/5/14	17.97	7.28	72	0.66	317
FW303	5/13/14	18.3	7.37	79	0.5	320
FW303	5/19/14	16.73	6.78	224	0.56	267.2
FW303	5/28/14	17.91	7.26	63	0.65	312.9
FW303	6/2/14	15.36	7.31	67	0.64	311.9
FW303	6/11/14	16.47	5.92	182	0.66	315.7
FW303	6/18/14	18.11	7.11	98	0.63	331.4
FW303	6/24/14	17.26	7.32	67	0.62	301.7
FW303	7/2/14	16.32	7.2	75	0.63	196.2
FW303	7/9/14	17.05	7.41	61	0.64	228.9
FW303	7/16/14	16.44	6.45	206	0.56	260.3
FW303	7/23/14	16.75	6.67	153	0.55	265.6
FW305	4/30/14	14.94	5.82	269	121	205.6
FW305	5/5/14	16.01	7.01	49	1.79	261.7
FW305	5/13/14	15.8	6.79	106	7.19	275.3
FW305	5/19/14	14.72	5.94	339	1.38	222.2
FW305	5/28/14	16.62	6.7	92	0.37	265.9
FW305	6/2/14	16.24	6.48	107	0.59	245
FW305	6/11/14	15.56	5.02	189	0.77	233.5
FW305	6/19/14	16.93	7.18	127	1.24	189.3
FW305	6/23/14	16.19	6.51	112	1.31	288.2
FW305	6/26/14	18.8	NA	277	1.55	281.1
FW305	7/7/14	18.3	NA	NA	2.833	274.8
FW305	7/9/14	16.42	6.26	144	2.2	21.01
FW305	7/14/14	22.8	7.2	NA	1.74	302
FW305	7/16/14	18	7.108	NA	1.38	275.1
FW305	7/23/14	16.65	5.56	184	2.11	92.31

bead that harbored the most or least diversity seemed to differ in an environment-dependent manner. For example, largest bead sizes from reactors were significantly more diverse than either small or medium sized beads. In contrast, the 425-600 μm sized beads harbored the greatest diversity from *in-situ* biosamplers incubated in FW301.

Larger Beads Tend to Have Greater Representation of Rare Populations

Across all bead sizes and all environments, the most abundant OTUs (with $>1\%$ abundance) represented a total of 29 bacterial classes (Fig. 4.5). Within any given bead sample, the most abundant populations were typically represented by less than 20 OTUs in each sample. The majority of OTUs were represented by either intermediate (relative abundance between 0.01% and 1%) and rare ($<0.01\%$) populations. Linear regression models were performed to assess the relationship between bead size and rarity. In beads incubated in FW303, the proportion of rare populations tends to increase directly with increase in bead size ($R^2=0.66$, $p\text{-value}=0.003$). This increase directly coincided with a decrease in representation from the intermediately abundant populations ($R^2=0.64$, $p\text{-value}=0.004$). This trend was not found in beads from the other two samples (FW301 and reactors), although the largest beads within reactors had significantly greater proportion of rare OTUs than the smallest bead size ($p\text{-value} = 0.046$).

Across all beads within reactors, populations of *Phenylobacterium* rank among the top OTUs in all bead sizes. In the smallest beads harvested from reactors, the top 20 OTUs comprised between 67 and 69% of total diversity (Fig 4.7). In addition to *Phenylobacterium*, highly abundant populations within smallest beads also included *Mesorhizobium*, *Novosphingobium*, and *Bradyrhizobiaceae*. In 425 μm beads, the top 20

OTUs make up to 75-79% of total community and are comprised of *Phenylobacterium*, *Novosphingobium*, *Chitiniphagaceae*, *Pedomicrobium*, and *Anaerolineae*. In 3000 μ m beads, top 20 OTUs make up 57-60% of the total community and are represented by *Pedomicrobium*, *Actinomycetales*, *Sediminibacterium*, and *Phenylobacterium*.

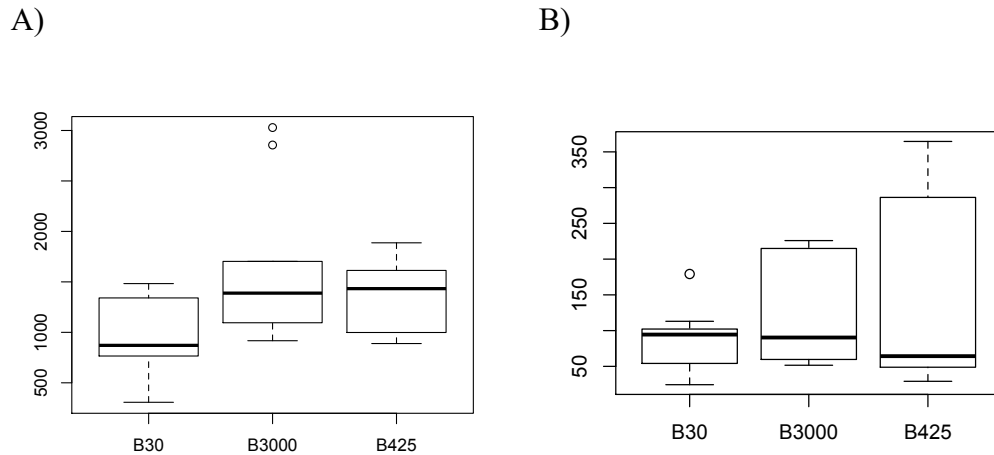


Figure 4.1 Boxplots of averaged A) Richness ($q=0$) and B) Hill Number ($q=1$) for each bead size across all environment types (reactors, FW301, and FW303). A one-way ANOVA was performed and tested with Tukey Honestly Significant Difference test (TukeyHSD). Richness was significantly different between smallest and largest beads (p -value = 0.011), although when 2 outlier points were removed (largest beads from reactor environment), the resulting difference is no longer significant. Diversity based on Hq1 was not significantly different for differently sized beads.

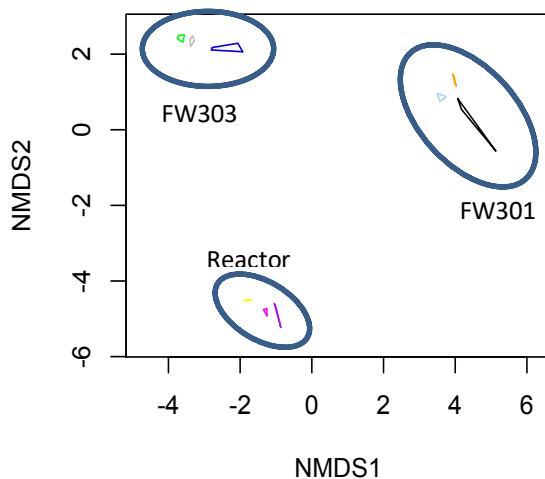


Figure 4.2 Representation of pairwise Bray-Curtis distances different sized beads incubated within Reactors, FW303, and FW301 and surrogate sediment samples using two dimensional NMDS ordination. Data converged after 20 random starts with stress value of 0.01.

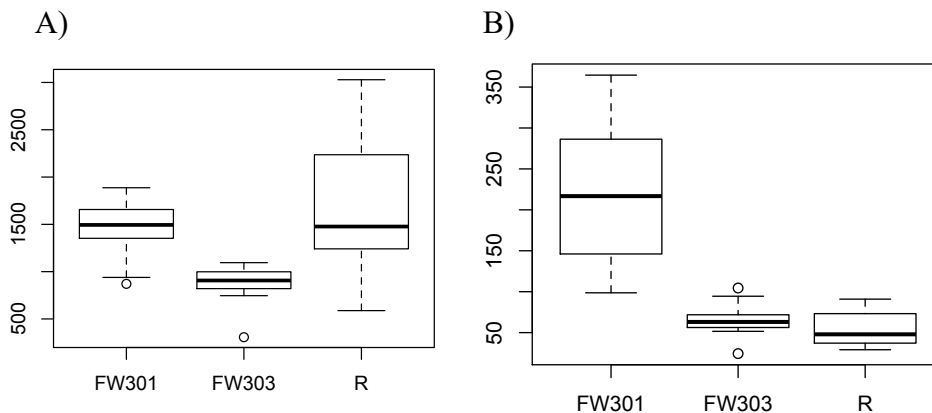


Figure 4.3. Boxplots of A) Richness and B) Diversity (Hq1) of bead samples across environment type. A) Richness observed on beads from FW303 was significantly lower than for beads from either FW301 or from laboratory reactors (R) (p -value = 0.014 and p -value=0.0019, respectively). B) Diversity observed on beads from FW301 was significantly higher than for those found in either FW303 or laboratory reactors (R) (p -value = 7×10^{-7} and p -value = 1.5×10^{-6} , respectively)

In 30-50 μ m beads of FW301 wells, the top 20 OTUs represent 43-56% of the total community and contain populations from *Phenylobacterium*, *Bradyrhizobium*, *Bradyrhizobialceae*, *Microbacterium*, and *Thiobacillus*. The intermediately sized beads harvested from this well display greater evenness and the top 20 OTUs comprising 31-34% of the total community, with *Chitiniophagaceae*, *Phenylobacterium*, *TM7*, *Bradyrhizobiaceae*, Candidate Phylum OD1, and *Mirococcaceae* ranking among the most abundant. Larger bead sizes contain OTUs of *Micrococcaceae*, Candidate Phylum OD1, *Thiobacillus*, *Mesorhizobium*, *Phenylobacterium*, *Micrococacceae*, most of which are each found between 5-8% relative abundance and cumulatively sum to 41% of the total community.

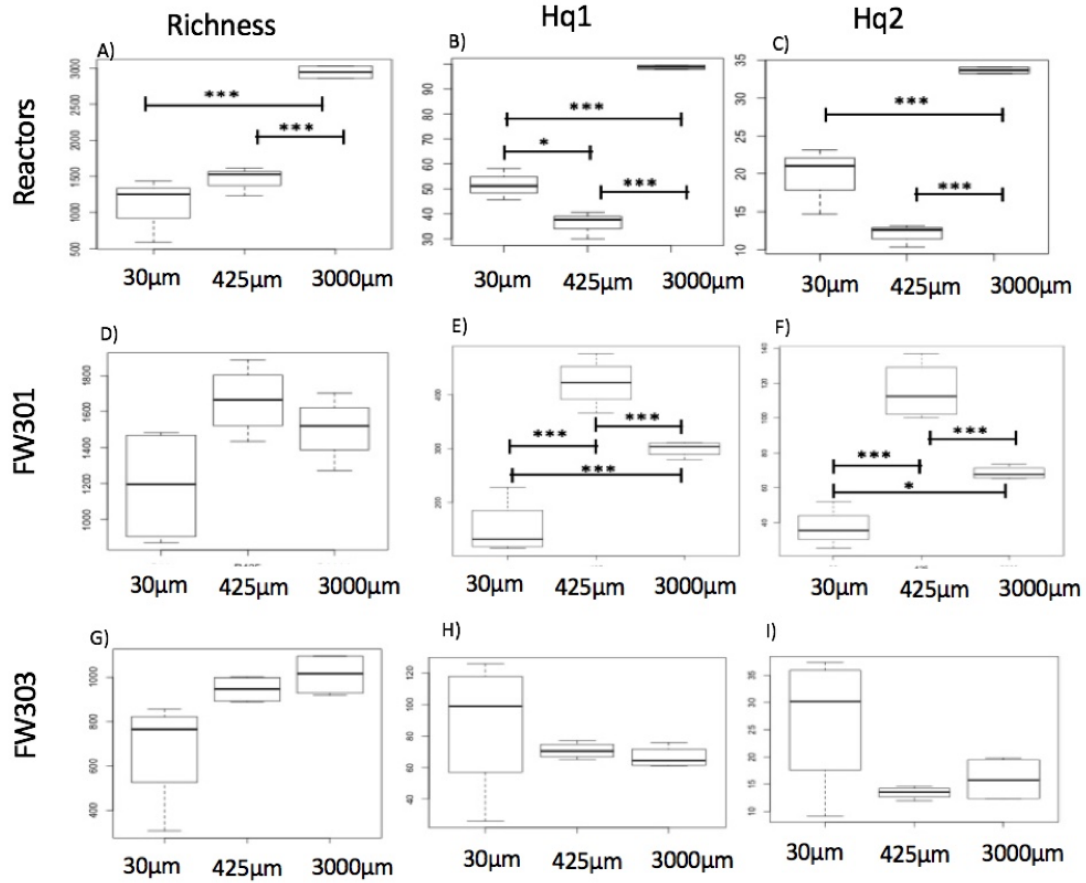


Figure 4.4. Comparisons of diversity measures based Hill Numbers of $q=0$ (Richness), $q=1$ (Exponential Shannon's Entropy) and $q=2$ (Inverse Simpson's) across bead sizes for each environment type (A-C, Reactors; D-F, FW301; G-I, FW303). A one-way ANOVA was performed and tested with Tukey Honestly Significant Difference test (TukeyHSD). Richness and diversity between beads from and FW301 reactors were significantly different ($p\text{-value} < 0.05$) as denoted with an (*). Richness and diversity between beads within FW303 were not significantly different ($p\text{-value} > 0.05$). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

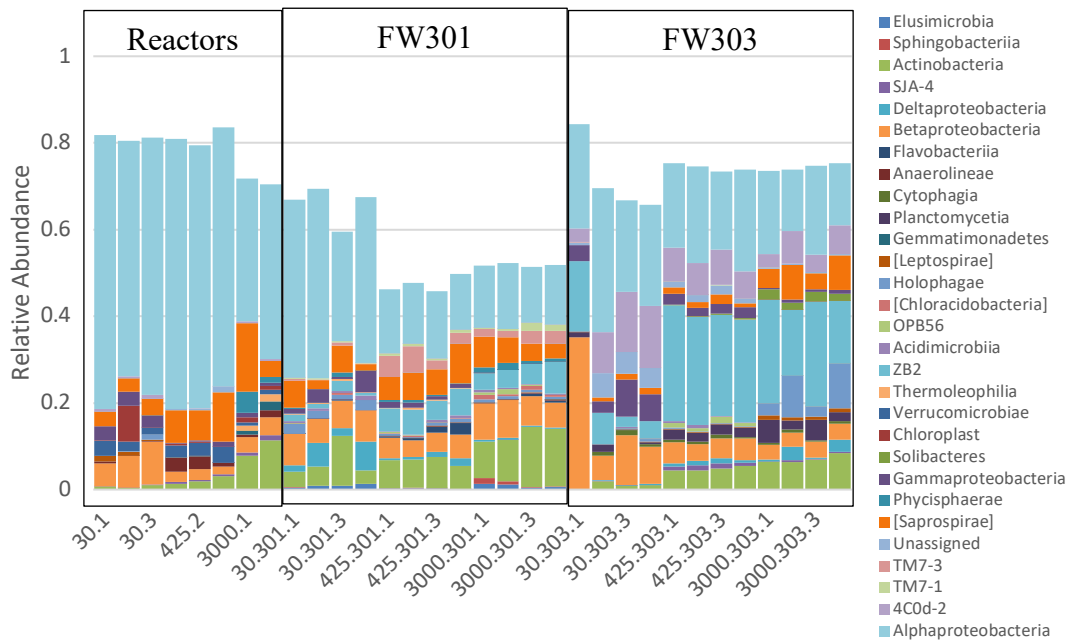


Figure 4.5. Relative abundances of populations by taxonomic class across all bead sizes (30 μ m, 425 μ m, 3000 μ m) and environment type (Reactors, FW301, FW303).

Across all beads of FW303, populations of Candidate Phylum OD1 are typically found among the most abundant OTUs, and dominate communities of intermediate and large bead sizes. Other abundant populations observed within the smallest beads of this well were *Oxalobacteracea*, *Cyanobacteria*, *Rhodospirillaceae*, *Caulobacteraceae*, *Gammaprotegocateria*, *Rhodospirillaceae*. Beads of 425-600 μ m size also included populations classified as *Rhodobacter*, *Cyanobacteria*, *Microbacterium*, *Caulobacteraceae*, while largest beads were observed to contain *Rhodobacter*, *Microbacterium*, *Holophagaceae*, and *Chitinophigaceae*.

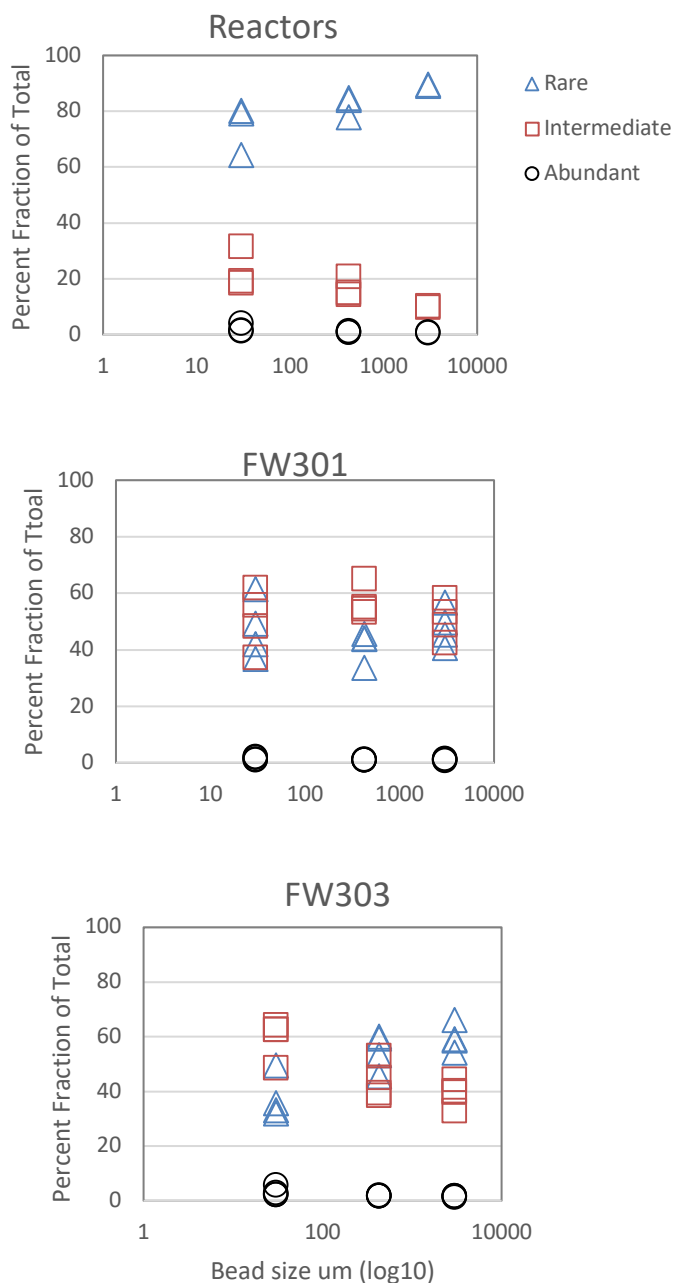


Fig 4.6. Percent fraction of OTUs residing on each bead sample per environment type whose relative proportions classified them as rare (blue), intermediate (red), or abundant (gray) populations. In FW303, rare populations (relative abundances <0.01%) comprised an increasing proportion of the total community as bead size increased ($R^2=0.66$, $p\text{-value}=0.003$). Conversely, contribution of intermediate populations (relative abundance between 0.01% and 1%) decreased as bead size increased ($R^2=0.64$, $p\text{-value}=0.004$). Similar trends were not observed in samples of FW301 or in reactors. However, communities of large beads in reactors also tended to have rarer OTUs than intermediate or small beads, and harbored significantly more rare populations than the smallest beads ($p\text{-value} = 0.046$)

Organic Matter Composition and Cellular Abundances Differ in Bulk Water of Laboratory Reactors Harboring Differently Sized Particles

Bulk fluid within each reactor was sampled for quantification of cellular abundance over the six-month incubation period. Even though all three reactors were inoculated with

the same source of groundwater, cell counts within the bulk fluid after 24 and 60 days of incubation showed greater cell abundances within the bulk (*e.g.* planktonic) fraction of reactors harboring large (3000 μ m) and intermediate (425 μ m) particles (Appendix C, Supplementary Fig. C.4). Additionally, bulk fluid of the smallest beads (30-50 μ m), source groundwater (Carboy) and non-bead control reactor (Water only) consistently showed lower cellular densities over the six-month period than did reactors that contained intermediate and larger sized particles. An increase in cellular abundance for all reactors and source groundwater was observed after about 60 days, coinciding with a pulse of humic acids injected directly into the source groundwater (carboy). Humic acids derived from sediments located in the background site of the ORR were added in an attempt to increase biomass within reactors in a field relevant manner. Interestingly, this time point is the only one for which water only (no bead control) showed higher cellular abundances that went above those of a reactor with particles (in this case, 30-50 μ m). After 120 days, cellular abundances within the bulk fluid of all reactors were more similar, and by the end of the six-month observation, all reactors harboring particles had bulk cellular densities ranging from $6.8 \times 10^3 \pm 870$ to $8.5 \times 10^3 \pm 345$ cells/mL, while carboy and no-bead control contained $4.6 \times 10^3 \pm 345$ cells/mL.

The bulk fluid of laboratory reactor effluent was also tested using EEMS analysis in order to detect and differentiate different species of organic matter that may be present within environments harboring microbial communities on differently sized particles (Appendix C, Supplementary Fig. C.5). EEMS analysis showed shifts in organic matter

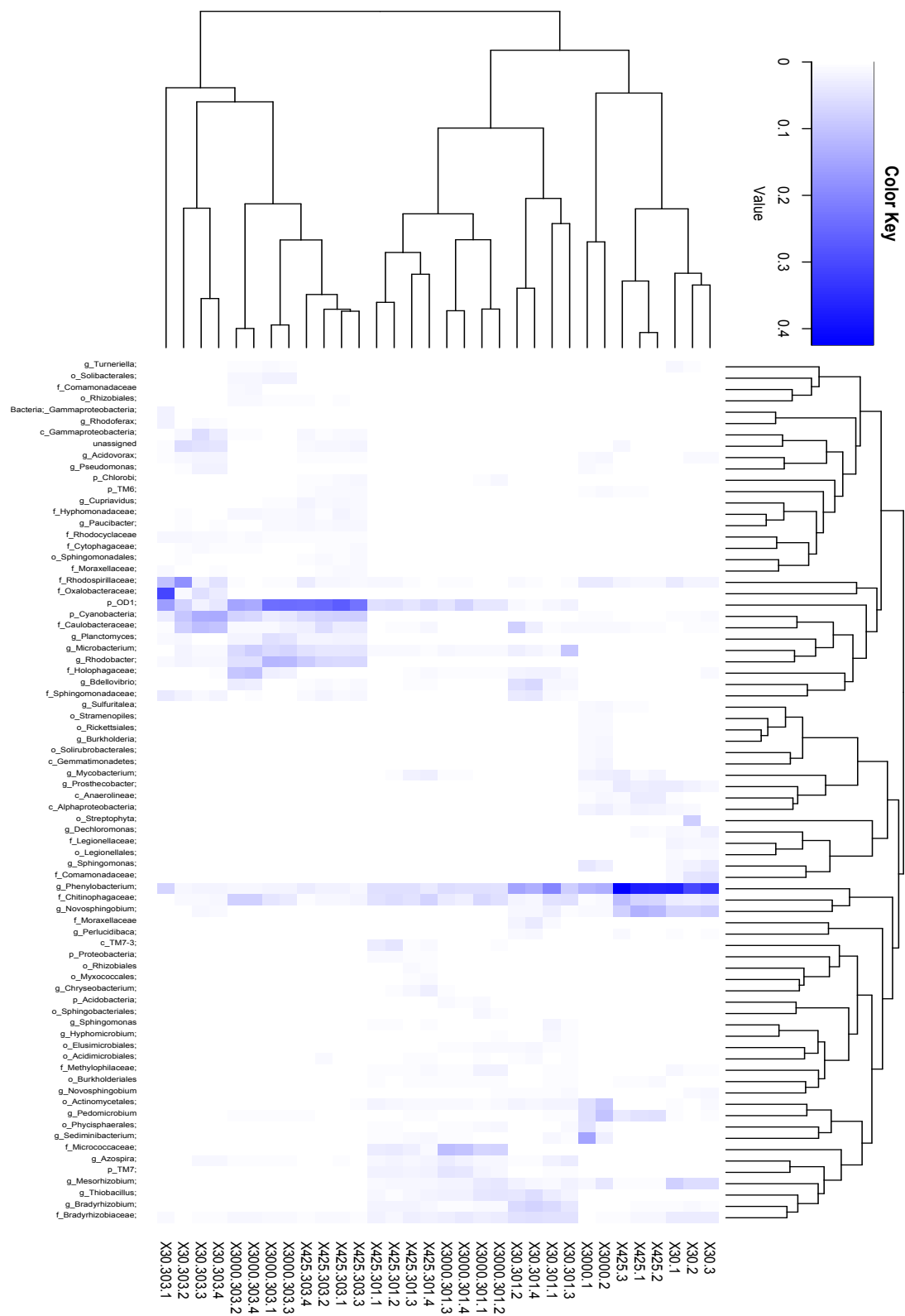


Figure 4.7 Heatmap of relative abundances of populations represented at >1% relative abundance across particle size (30µm, 425 µm, 3000 µm) and environment. Samples were clustered using the average linkage method based on similarity of community composition across samples. Taxa were clustered based on their co-occurrence across samples.

composition, with a decrease in more labile, protein-like organic acids in the effluent of reactors coinciding with increasing bead size (Appendix C, Supplementary Fig. C.5). Additionally, there was an increase in more recalcitrant, humic-like organic acids in reactor effluent that coincided with increasing bead size.

Discussion

Micro-niches and microenvironments that exist within porous soils and sediments of the subsurface are known to vary considerably in several parameters important to microbial life. In this study, sterile and previously uncolonized glass beads were incubated *in-situ* and in laboratory conditions in order to investigate the impact of physical surface scale on microbial community structure.

Overall, community structure was more greatly influenced by the source microbes of the surrounding environment than by bead size alone (Fig. 4.1 and Fig. 4.3). Laboratory reactors were fed with unfiltered groundwater from FW301, which also served as one of the test field wells in which surrogate bead samplers were incubated. Despite laboratory reactors receiving representative FW301 groundwater, samples from FW301-fed reactors and FW301 surrogate bead samples were mostly distinct and unique (Fig 4.2), likely due to differences in hydrodynamics between the two environments. Beads from laboratory reactors had a much greater proportional representation of α -proteobacteria, and less

representation of Actinobacteria than did beads incubated in FW301. It is possible that source microbes captured in the 20L of FW301 formation groundwater that was used to feed laboratory reactors resulted in a bottle-neck effect which limited the resulting local bead community structure. Laboratory beads were also maintained in a continuously mixed system, which may not reflect the *in-situ* heterogeneities in terms of nutrient flux and groundwater flow that has been previously shown to have significant effects on daily richness in FW301 (Zelaya et al., submitted). Ideally, beads incubated *in-situ* would have been directly compared to surrogate sediments incubated in the same field wells. Attempts were made to perform a meta-analysis of the two datasets, however, technical differences between the methods used for extraction, amplification (including the targeting of two different regions of the SSU rRNA gene, sequencing, and processing of the samples rendered such a comparison unsuccessful. Currently, bead samples collected from *in-situ* surrogate sediments are being processed using identical methods as those used for surrogate sediment samplers so that direct comparisons of particle-size effects on community structure resulting from *in-situ* conditions can be better discerned.

Results from a previous study that used identical reactor set up on pure cultures of a field-relevant *Desulfovibrio* isolate showed that biofilm protein per surface area ($\mu\text{g}/\text{cm}^2$) was 25-fold and 50-fold greater for the intermediate and largest sized particles, respectively, compared to the smallest (Altenburg et al., unpublished results). However, the overall biofilm carbohydrate to protein ratio was similar for all bead sizes (0.11, 0.07, 0.09, respectively). Additionally, it was observed that the amount of biofilm per unit area decreased with the particle size even as surface area/volume increased. One possible

explanation for the counter-intuitive result is that the growth of biofilm significantly alters the porous structure and consequently changes porosity, permeability and dispersivity of the substratum and that the effect was more pronounced for smaller particles.

Previous studies into the community structure across various aggregate sizes have shown that larger aggregates ($>2\text{mm}$) can support more diverse microbial communities than smaller microaggregates ($<2\text{mm}$) (Davinic et al., 2012). While total richness and diversity was not directly related to particle size alone in the present study, larger particles did tend to harbor a greater proportion of rarely abundant populations. While rarity was arbitrarily defined as populations that represented less than 0.01% of the total community in this study, this observation only held true for two of the three sampled environments (laboratory reactors and FW303) (Fig. 4.6). The role of mixing in aquatic systems has been implicated in the creation of patchy distributions of both nutrients and biomass (Ebrahimi and Or 2016; 2018). Additionally, particle aggregation induced by microorganisms has been shown to alter pore structure and create preferential flow paths that allow for unique microbe:microbe ratios (Crawford et al., 2011). Interestingly, FW303 has previously been described as a more stable and homogenous field well in terms of geochemical variability and daily groundwater community structure as compared to FW301 (Zelaya et al., submitted). It is possible that the hydrogeological and microbial dynamics in FW301 resulted in unique behavior in terms of source populations, biofilm formation, and subsequent porosity, permeability and cellular dispersal at the particle scale not observed in the other two environments.

CHAPTER FIVE

CONTRIBUTION OF TOTAL AND VIABLE DNA POOLS ON DIVERSITY
ESTIMATES OF CONTAMINATED AND NON-CONTAMINATED SUBSURFACE
SEDIMENTS AND GROUNDWATER

Contribution of Authors and Co-Authors

Manuscript in Chapter Five

Author: Anna J. Zelaya

Contributions: Performed literature review, synthesized ideas, performed statistical and bioinformatical analyses, wrote and edited the manuscript

Co-Author: Heidi J. Smith

Contributions: Synthesized ideas, performed statistical and bioinformatical analysis, edited the manuscript

Co-Author: Frederick von Netzer

Contributions: Field sampling and data collection

Co-Author: Matthew W. Fields

Contributions: Principal Investigator, Synthesized ideas, wrote and edited the manuscript

Manuscript Information Page

Anna J. Zelaya, Heidi J. Smith, Frederick von Netzer, Matthew W. Fields

Status of Manuscript:

☒ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☐ Published in a peer-reviewed journal

ABSTRACT

Nucleic acids originating from lysed or otherwise dead cellular sources have been detected to persist in natural environments. Extracellular or “relic” DNA originating from non-living microbial cells has the potential to skew microbial diversity estimates based on sequencing surveys. In groundwater and sediment fractions of the subsurface environment, the potential effect of DNA from dead or dying microbial cells has not been fully assessed. Microbial communities are the main drivers of geochemical processing in subsurface environments; however it remains unclear which specific populations within the subsurface are both intact/viable and active. In an attempt to delineate viable/intact cells and active populations within subsurface environments, unfiltered groundwater and freshly cored sediment from both contaminated and non-contaminated environments were subjected to three different but complementary methods, 1) Bioorthogonal non-canonical amino acid tagging (BONCAT), 2) ^3H -Leucine incorporation, and 3) Propidium Monoazide Sequencing (PMA-Seq). As this study was part of a larger collaboration between multiple teams and institutions, this chapter will focus on results from the third method, PMA-Seq, as it was the primary contribution of the author. Unfiltered groundwater was either treated or untreated with PMA and subsequently filtered through 0.2 μm and 0.1 μm filters to capture microbial biomass for 16S rRNA sequencing. For all groundwater communities captured on 0.1 μm filters, OTU richness and diversity were similar with and without PMA treatment, despite significant differences in amplifiable 16S rRNA gene copy number (p -value <0.05). However, in the 0.2 μm fraction of non-contaminated groundwater, PMA treatment resulted in greater diversity of dominant populations (p -value=0.04). This result

suggests that DNA from potentially non-viable members of dominant populations may influence sampled community estimates in non-contaminated groundwater. Uncontaminated cored sediments showed distinct community profiles across depth, and DNA from non-viable members may have a decreasing influence on sampled community diversity as depth increases. Overall, these results suggest that the majority of microbial populations detected in molecular surveys of subsurface groundwater and sediments are largely representative of intact/viable cells. These studies, in combination with other methods that further distinguish between activity rates and physiological states, can be applied towards the increased understanding of cellular status and physiological states of subsurface microorganisms across spatial and temporal scales.

Introduction

Despite low amounts of organic carbon, nutrients, and electron acceptors (Jones et al. 2018), the terrestrial subsurface harbors a vast amount of microbial biomass and diversity (Hug et al., 2016). Additionally, it has been estimated that terrestrial subsurface environments contain upwards of 40% of the microbial biomass on Earth (Whitman, Coleman and Weibe, 1998; Griebler and Leuders, 2009; McMahon and Parnell, 2014). Over the last 20 years, the use of DNA-based approaches have increased our ability to study the wide diversity of mostly uncultivated microorganisms (Bernard et al., 2018). As a result, our understanding of more deeply sampled microbial composition, diversity, and distribution across a wide range of environments has improved (Amman and Rossello-Mora, 2016; Lennon and Locey, 2016).

Due to limitations in statistical and molecular methods, most microbial diversity surveys within a given environment have historically focused on the most abundant populations (based on genetic markers such as SSU rRNA gene sequence abundances) (Galand et al., 2009; Pedros-Alio, 2012, Jousset et al, 2017 and references therein). Emphasis on the dominant fraction of a community is understandable given the limited quantitative data provided lowly abundant, sparsely present, or absent populations (Haegeman et al., 2013; Weiss et al., 2016). However, it has been recently argued that rare populations deserve greater attention. Recent studies have shown that rare members could serve as a microbial “seed bank” that could prevent ecosystem collapse in changing environmental conditions (Shade et al., 2014; Lynch and Neufeld, 2015; Jousset et al., 2017). In addition, rare populations of microbial communities have been shown to be physiologically active and play over-proportional roles in biogeochemical cycles (Sogin et al., 2006; Jousset et al., 2017).

In natural systems, microorganisms can be observed in various physiological states (*e.g.* active, lowered metabolism, dormant, various stages of death), and linking microbial identity with functional activity is a central goal in microbial ecology. Typical DNA extraction methods for molecular surveys are often insufficient to discriminate between DNA pools and the physiological state of the source cell. Additionally, the persistence of DNA within the environment (Levy-Booth et al, 2007; Pietramellara et al., 2008) may obscure DNA-based estimates of microbial diversity (Carini et al., 2016). Recent advances in molecular techniques have provided an assortment of methods available to assess active and viable fractions of communities. The presence and quantitation of short-lived RNA

molecules can be used to infer actively-growing cells (Hirsch et al., 2010). The incorporation of amino-acid analogues, such as in bioorthogonal non-canonical amino-acid tagging (BONCAT) (Hatzenpichler et al., 2015), stable isotopes (Lueders *et al.* 2016, Smith et al. 2017), and radioisotopes (Cottrell and Kirchman, 2000) can distinguish metabolically active members within environmental samples. Moreover, the use of selectively permeable dyes (*e.g.* propidium monoazide) can help to differentiate between viable and non-viable populations based on cellular membrane integrity (Knocker et al., 2006).

Propidium monoazide (PMA) is a photo-reactive dye that binds to double-stranded DNA (Knocker et al., 2006) and selectively permeates cells with compromised cellular membranes, enabling for the differentiation between viable (membrane intact) and non-viable (membrane compromised) cells (Fig. 5.2). PMA has been used to distinguish between pools of DNA in environmental samples such as Antarctic lake ice (Dieser et al., 2010), surface soils (Carini et al., 2016), and deep marine sediment (Ramirez et al., 2018). Recently, the use of PMA demonstrated that abundance estimates for dominant soil microorganisms were overinflated by “relic” DNA, resulting in the underestimation of more rare community members (Carini et al., 2016). However, these results may be environment specific, as similar studies in municipal wastewater (Knocker et al., 2007), Antarctic lake ice (Dieser et al., 2010), and deep marine sediments (Ramirez et al., 2018) showed minimal influence of extracellular DNA on diversity estimates.

Within groundwater of a shallow subsurface aquifer it was previously observed that microbial richness and diversity varied significantly over short-time scales (Zelaya et

al., submitted). Results indicated that groundwater habitats may be populated by transiently rare and transiently dominant populations over short time scales. Additionally, comparisons between groundwater and surrogate sediment samples showed that groundwater communities had greater overall richness and diversity than those of corresponding surrogate sediment. Relationships between observed changes in diversity (*i.e.* species richness) and possible abiotic drivers (*i.e.* geochemistry) were mostly weak for the tested time period, with few exceptions. Therefore, the observed transient nature of groundwater populations remains mostly unexplained. It is possible that variation may be due to biotic a combination of biotic and abiotic factors over time. While models have been developed to quantify the relative contributions of biotic drivers of community assembly (*e.g.* β MNTD and β NTI), such models still rely on sampled SSU rRNA gene abundance. Therefore, unless a thorough understanding of the extent to which the SSU rDNA fraction derives from living/dead/dying sources within natural environments, the potential for misleading diversity estimates will continue.

In order to assess the extent to which bacterial populations in groundwater and sediments are comprised of membrane intact versus membrane compromised cells, and the influence of increased environmental stress (*e.g.* contamination) on cellular integrity, both planktonic and sediment-associated bacteria from contaminated and non-contaminated environments were investigated. This study was part of a larger collaboration between multiple teams and institutions, with the ultimate goal of assessing cellular integrity and activity of subsurface microorganisms. However, this chapter will focus on results from the third method, PMA-Seq, as it was the primary contribution of the author.

Materials and Methods

Site Description and Sample Collection

The study site is located in the Y-12 National Security Complex in Bear Creek Valley (BCV) in Oak Ridge, Tennessee, USA. The study obtained samples from within the vicinity of the former S-3 ponds, which contain contaminated waste associated with the refinement of uranium ore (Fig. 5.1). The geological and hydrological setting within the vicinity of the former S-3 ponds has been previously described (Watson et al., 2004). In brief, bedrock depth ranges between 5–10 meters below ground surface (mbgs). The depth to groundwater is ~2.5–3 mbgs and the water table level is largely controlled by precipitation and infiltration, varying up to 1 meter during major precipitation or drought events. Groundwater flow within the unconsolidated materials is largely controlled by ground surface elevation and is generally south to southwest and towards Bear Creek. Hydrological flow regimes have been difficult to estimate within the subsurface, although estimates of hydraulic conductivity (K) vary by up multiple orders of magnitude, and subsequent residence time for the background area are estimated to range from 3 hours to 30 days (Watson, et al., 2004; Zelaya et al., submitted).

Sediment collection

Freshly cored samples of unconsolidated materials were obtained from boreholes EB271 and EB106 (Fig. 5.1). Core sample EB271 is comprised of uncontaminated sediments as was sampled from an area located up-gradient of the former S-3 ponds, whereas core sample EB106 is comprised of contaminated sediments as was sampled from an area located down-gradient and as such is contaminated with uranium and nitric acid

whereas EB271 was not. The boreholes were advanced using a dual tube (DT22) direct-push Geoprobe drill rig resulting in a sealed casing through which undisturbed sediment samples are recovered. Sediment samples were encased in disposable thin-walled polyvinyl chloride (PVC) liners (152.4 cm length x 2.86 cm I.D.) attached to 3.18 cm-

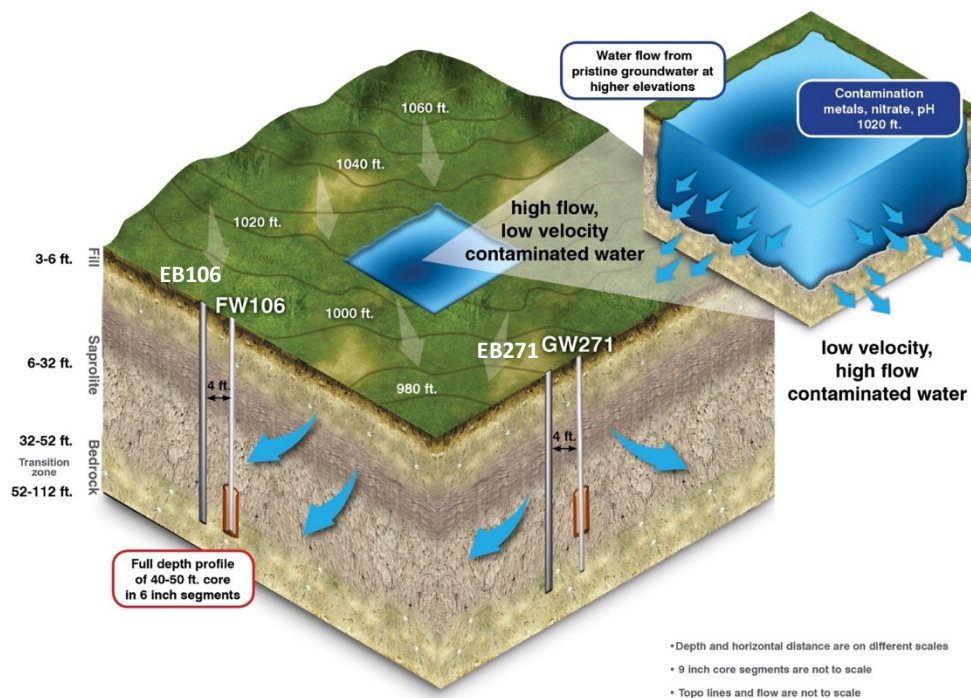


Figure 5.1. Cross sectional view of sampled sediment cores (EB271 and EB106) and groundwater wells (GW271 and FW106) in proximity to S3 waste ponds at the EOR-FRC. Credit: Joseph Parschen, Center for Biofilm Engineering, Montana State University-Bozeman.

outside diameter inner rods (Geoprobe, 2011). Core segments (15.24 cm) were immediately capped and stored under a nitrogen atmosphere at 4°C. Cores were aseptically processed within 24 hours of collection with aseptic techniques as follows: core segments were transferred into an anaerobic chamber (5% H₂/95% N₂; Coy Laboratories, Ann Arbor,

MI), end caps were removed and intact subcore samples were extracted from each end with an autoclaved syringe barrel with the flange removed. Sediment along the core length was exposed by longitudinal cuts through the PVC liner with a liner cutter (DT32/DT325, Geoprobe). Samples from three core-depths representing unsaturated (137.16cm bgs), transition or capillary fringe (220.97cm bgs), and saturated (232.74cm gbs) zones were selected for study using a variety of complementary methods, for which PMA-seq will be discussed. At present, only results for the non-contaminated core will be presented in this chapter.

Groundwater Collection

Groundwater samples were collected from neighboring monitoring wells, GW271 for EB271 and FW106 for EB106, respectively (Fig. 5.1) by low-flow purge and sampling. Groundwater was purged with a peristaltic pump connected to dedicated down-well tubing installed at the mid-screen of the wells. A water quality indicator probe (Troll 9500, In-Situ Inc.) was connected in-line to a flow through cell to monitor pH, conductivity, and oxidation-reduction values during purging. Once these parameters stabilized, samples were collected in an appropriate aseptic/autoclaved container, preserved, and transported overnight at 4°C for laboratory analysis. All subsequent analyses (*e.g.* PMA) were performed within 24 hours of collection.

Propidium Monoazide Treatment For Groundwater and Sediment

The viability dye, PMAxxTM, used in this study is a modified version of propidium monoazide (Biotium, catalogue number #40069). According to the manufacturer,

PMAxx™ better eliminates the potential for false positives (incomplete elimination of dead cell DNA) that was a limitation of the original PMA formula (Biotium website). PMAxx™ will here-to-fore be referred to as PMA for simplicity.

Groundwater samples were processed as follows: 1L of groundwater from either FW106 or GW271 was stored in duplicate sterile 1L glass bottles (Total 2L per well). PMA was added to one of the 1L bottles at a final concentration of 10µM, while the second liter served as a non-treated control. Both PMA and non-PMA treated samples were incubated in the dark with gentle rotation (Model 55 rocking shaker, setting 20, Midwest Scientific) for 30 minutes at room temperature. After incubation, PMA-treated groundwater was aseptically transferred to a sterile glass tray (volume capacity of 1.9L), and exposed to a 650-watt halogen lamp placed 20 cm from the sample surface for four consecutive 30s:30s light:dark cycles in order photo-deactivate unbound PMA. Throughout photo-deactivation, groundwater was gently and continuously mixed (LabNet Orbit P4, 150rpm) to ensure even light exposure. Additionally, to prevent overheating due to light exposure, the glass tray containing groundwater sample was placed in a secondary container filled with ice. After photo-deactivation, 500mL was filtered through a 0.2µm filter, resulting in two sub-samples per treatment per well. The flow-through from the 0.2µm filters was collected and further filtered through a 0.1µm filter. Non-PMA treated groundwater was transferred to a different sterilized glass tray and processed identically as PMA treated samples. Filters were immediately frozen at -20°C until DNA extraction.

For cored sediment, 12 replicate sub-samples of each chosen depth (unsaturated, saturated, and transition) were resuspended (1% w/v) in sterile PBS (pH 7.4) in transparent

tubes. PMA was added to a final concentration of 40 μ M into six of the sub-samples, while the other six served as non-treated controls. Samples were incubated for 30 minutes as described for groundwater. After incubation, tubes were placed on ice and exposed to light under gentle rotation for the photo-deactivation step as described for groundwater. After photo-deactivation, tubes were immediately frozen at -20°C until DNA extraction.

DNA Extraction, and SSU rRNA Gene Amplification

Both groundwater and cored sediment were extracted using MoBio PowerSoil DNA extraction kit, following manufacturer's instructions. For groundwater, each filter was cut into four equal quadrants and DNA was extracted from each quadrant. Dividing each filter into fourths allowed for a maximum of 8-technical groundwater filter replicates per L. For cored sediment, DNA was extracted from 750 μ L (triplicate aliquots of 250 μ L) from each replicate sample containing sediment:PBS slurry.

Both groundwater and cored sediment samples were PCR amplified in triplicate 25 μ L reactions containing 0.1 μ M of each Illumina-barcoded primer, a 1X final concentration of KAPA HIFI HotStart Master Mix (Kapa Biosystems), and 2.5 μ L of 1X Bovine Serum Albumin (BSA). The primer design consisted of Illumina adaptor sequences followed by universal forward 341F 5'-tcgtcggcagcgtcagatgtgtataagagacagCCTACGGGNBGCASCAG-3' or 805R 5'-gtctcgtgggctcggagatgtgtataagagacagGACTACNVGGGTATCTAATCC-3' primers. PCR was performed in an Eppendorf Mastercycler Pro S. The amplification protocol consisted of an initial denaturation at 98°C for 3 min, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 15 s, extension at 68°C for 30 s, and a final extension step

at 72°C for 1 min. Negative PCR controls without DNA template were amplified to monitor contamination. There was no evidence of amplification from any of the tested procedural blanks and negative sorts and these samples were therefore not included in successive analyses.

Sequencing and Sequence Analysis

SSU rRNA gene amplicons were sequenced on an Illumina MiSeq Platform (San Diego, CA, USA) following the manufacturer's Metagenomic 16S Sequencing protocol for 300-read paired end sequencing. Following endpoint PCR amplification, amplicons (containing MiSeq adapter sequences) were purified (AmpureXP beads) and indexed to ensure accurate identification of individual samples following pooling and sequencing. Indexed amplicons were then bead purified, quantified with PicoGreen (Quant-IT, Invitrogen, Carlsbad, CA, USA), and concentrations normalized. Following the normalization of DNA concentrations samples were all pooled together, mixed with the PhiX control (10% of sample DNA concentration), and sequenced.

Forward and reverse sequences were joined with the Quantitative Insights Into Microbial Ecology (QIIME) toolkit version 1.9.0 (Caporaso *et al.* 2010). Within QIIME, sequences were quality refined, and sequences containing ambiguous bases, homopolymers >8 bases, an average quality score below 30 over a 50 bp window were excluded from further analysis. Operational taxonomic units (OTUs) were picked using the open reference picking method within QIIME and sequences were clustered to 97% similarity. Chimeric sequences were removed using UCHIME (Edgar *et al.* 2011) with the SILVA Gold database (Pruesse *et al.* 2007) and a second chimera check using sequences

from the present study as a database. Taxonomic assignments were made blasting representative OTU sequences to the NCBI database. OTUs with sequences less than $n=2$ across all samples were removed from the dataset. The result was an OTU table without singletons.

Dissimilarities between the samples based on the relative abundances of the top 1% populations were assessed via Sorensen and Bray-Curtis pair-wise dissimilarities using VEGAN in R statistical package (Oksanen, et. al, 2013). Unconstrained ordinations were carried out in R statistical package with VEGAN (Oksanen, et. al., 2013). Clustered heatmaps were created using R statistical package based on the relative abundances of the sampled groundwater for OTUs that represented the >5% of the relative abundance in at least one sample.

Replicate samples were grouped together and the most abundant OTUs (greater than 1% of the entire sequence library) were subjected to similarity percentage (SIMPER) analysis using PAST version 3.10 (Hammer *et al.* 2001). Samples were clustered using the average linkage method based on similarity of community composition across samples. Taxa were clustered based on their co-occurrence across samples. Replicate samples were grouped together and the most abundant OTUs (greater than 1% of the entire sequence library) were subjected to similarity percentage (SIMPER) analysis using PAST version 3.10 (Hammer *et al.* 2001).

qPCR

Quantitative PCR of the bacterial SSU rRNA gene was performed using primers 515F(5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-

GGACTACNVGGGTWTCTAAT -3') using an ABI StepOne Plus Real Time PCR System. Triplicate 25 μ L reactions were performed for all samples. Each 25 μ L reaction contained AB PowerSYBR Green Master Mix (12.5 μ L), Forward (0.5 μ L) and Reverse (1 μ L) primers, template or standard (2.5 μ L), and nuclease free molecular grade water (8.5 μ L). Reactions were performed under the following thermocycler conditions: 95°C (10m), 35x[95°C (15s), 54°C (60s), 72°C(15s)]. Additionally, a melt curve was performed using the following settings: 95°C (15s), 60°C (1m), 95°C (15s). gBlocks Gene Fragments (Integrated DNA Technologies, Inc.) were constructed from *E. coli* with 515F-806R primer set and were used as DNA standards. Standard concentrations used for each run ranged from 10⁴ to 10⁶ 16S rRNA gene copy numbers per microliter.

Rationale for PMA

Several techniques have been developed to investigate microbial viability and activity in a variety of environments (Nocker 2006, Hatzenpichler 2016) however, there are known limitations associated with each methodology. PMA was used to distinguish between different DNA pools, namely i) cells with intact cellular membrane integrity, ii) cells with compromised cellular membrane integrity, and iii) extracellular DNA. While rRNA-based approaches are thought to discriminate between metabolically active and inactive cells, recently this notion has been challenged (Blazewicz et al. 2013). These approaches are further complicated by low rRNA yields from soil habitats (particularly from clay rich soils). While there is the potential for differences in cellular susceptibility to PMA treatments, recent studies have demonstrated that there are limited taxon specific differences in the susceptibility of intact cells to PMA treatment (Carini et al. 2016).

Another potential limitation is the adsorption of PMA to soil or charged particles, which inhibits binding to relic DNA and could impact photo-deactivation (Fittipaldi et al. 2012). However, to address this potential limitation, representative groundwater was spiked with DNA and treated with PMA to ensure total removal with the concentrations used in the present study (described below).

Rational for PMA Concentrations Used

Environmental samples from this study contained approximately 10^4 - 10^5 cells/mL (groundwater) and approximately 10^6 cells/gram (sediment) (data not shown). Two methods were used to ensure that appropriate PMA concentrations were used to eliminate DNA originating from non-viable cells in groundwater. First, *Escherichia coli* was grown in LB medium to mid-exponential phase and serially diluted to ranges from 10^5 to 10^7 CFU/mL (16 total replicates per concentration). Eight tubes for each CFU/mL concentration were chosen to serve as killed controls (heated at 72°C for 10 min). For both live and killed sets, four tubes from each set were randomly selected for PMA treatment at 10 μ M concentrations, resulting in four replicates each of i) Live, PMA treated ii) Killed, PMA treated, iii) Live, PMA untreated and iii) Killed, untreated. Tubes were incubated and photo-deactivated as described above, followed by immediate DNA extraction and amplification. Results are available as supplementary material (Appendix D, Supplementary Fig. D.1.A), and demonstrate that i) PMA did not enter living cells or reduce amplifiable DNA in living cells, and ii) 10 μ M was enough to eliminate up to 10^7 CFU/mL of killed *E. coli* grown in pure culture.

In order to ensure that charged particles within the groundwater environment did not adsorb PMA and prevent PMA-DNA binding, a second approach to ensure appropriate PMA concentration for groundwater was performed. In this second approach, eight clear screw cap tubes with 1mL of representative non-contaminated groundwater was spiked with either 47ng or 470ng of purified DNA extracted from *E. coli*. This DNA quantity corresponds to $\sim 10^6$ and $\sim 10^7$ bacterium chromosomal equivalents, assuming 4.7 fg of DNA per bacterial cell (Button and Robertson, 2001). From these eight tubes, four were randomly selected for PMA incorporation to a final concentration of 10 μ M. The remaining four tubes acted as non-treated controls. Samples were processed as previously described. Following DNA extraction, SSU rRNA genes were quantified with real-time PCR. The results can be found in Appendix D (Supplementary Fig D.1.B), and demonstrate that the concentration used in this study (10 μ M) was sufficient to remove DNA representing up to 10^6 bacterial cells from representative groundwater.

Due to cost constraints, a similar study was not performed to test that 40 μ M of PMA was sufficient for sediment samples. Instead, the rationale for the concentration used for sediment was based on the method following Carini et al., 2016, in which 40 μ M of PMA was sufficient to remove spiked DNA for a variety of soil types (Carini et al., 2016).

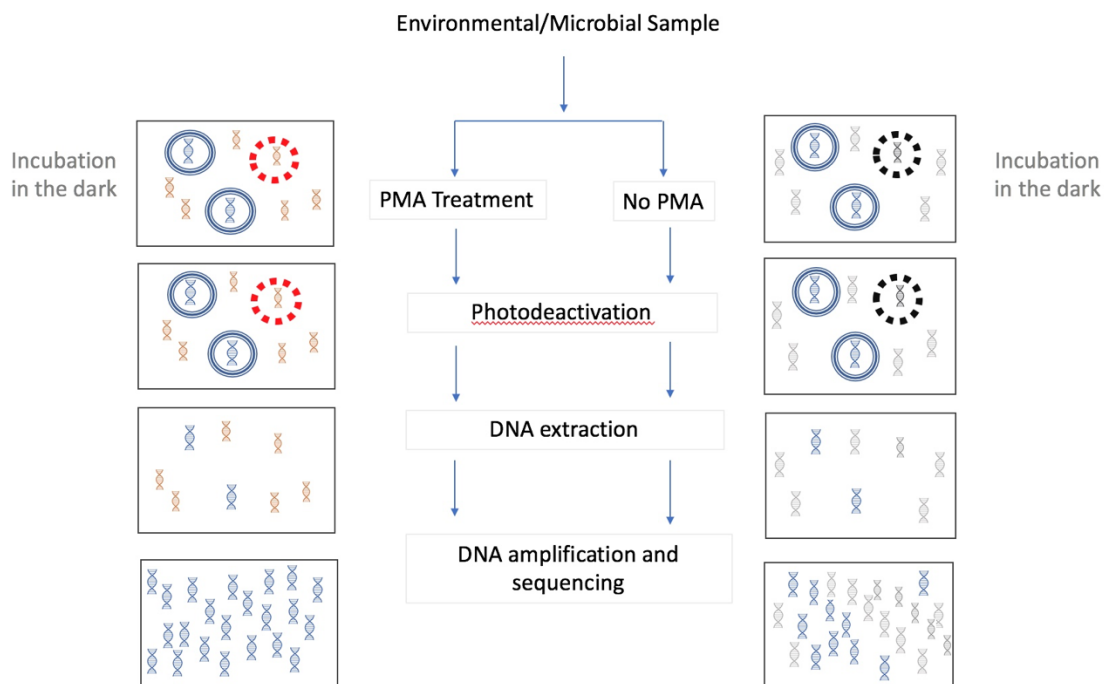


Figure 5.2. Viability PCR workflow using photoreactive and selectively permeable DNA intercalating dyes (e.g. PMA). Different DNA pools [*i.e.* eDNA (grey), dead cells (black), living cells (blue)] are distinguished via selective PMA-DNA (red) binding and photoactivation. A non-treated samples serves as a negative control and can reveal relative proportions of viable/non-viable cells.

Results

PMA Has Minimal Influence on Total Richness and Diversity of Contaminated and Non-Contaminated Groundwater

It was expected that PMA would either have no effect (*i.e.* similar richness would be observed between treated and non-treated samples) or PMA would remove any portion of the DNA pool that originated from extracellular or non-viable sources, resulting in significantly different richness and diversity. While PMA treated samples showed higher overall richness and diversity than non-treated samples (Fig. 5.3C,D) for the majority of samples, these differences were mostly statistically insignificant ($p\text{-value} > 0.05$). As such,

PMA had minimal influence on overall community structure in terms of richness and diversity for most groundwater samples. However, in non-contaminated groundwater, diversity measures that downweigh the influence of rare members (Inverse Simpson) showed that the PMA treated 0.2µm fraction was statistically more diverse than the non-PMA treated fraction ($p\text{-value} = 0.04$), indicating that the dominant community structure was influenced by the presence of DNA from potentially non-viable sources.

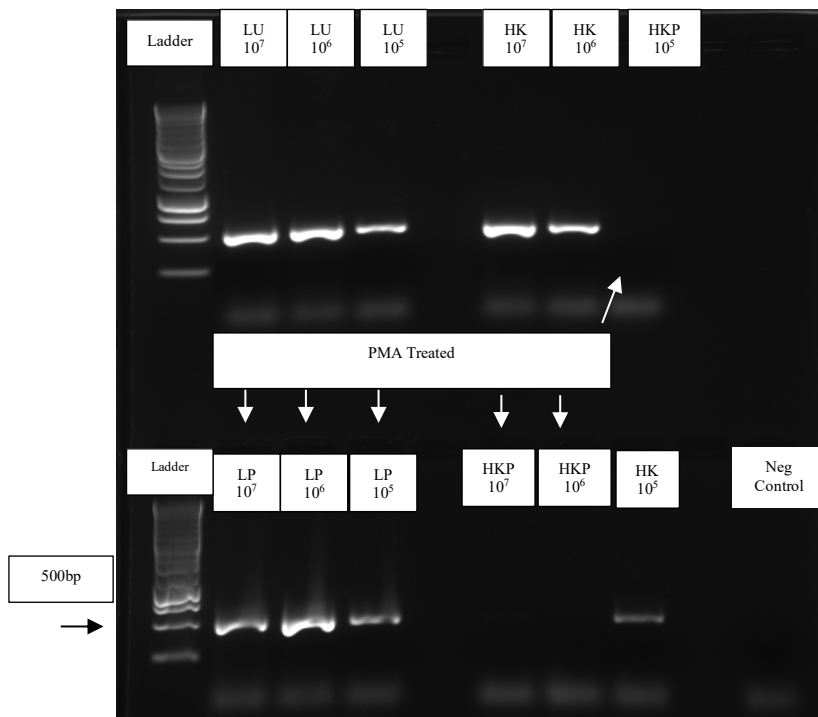
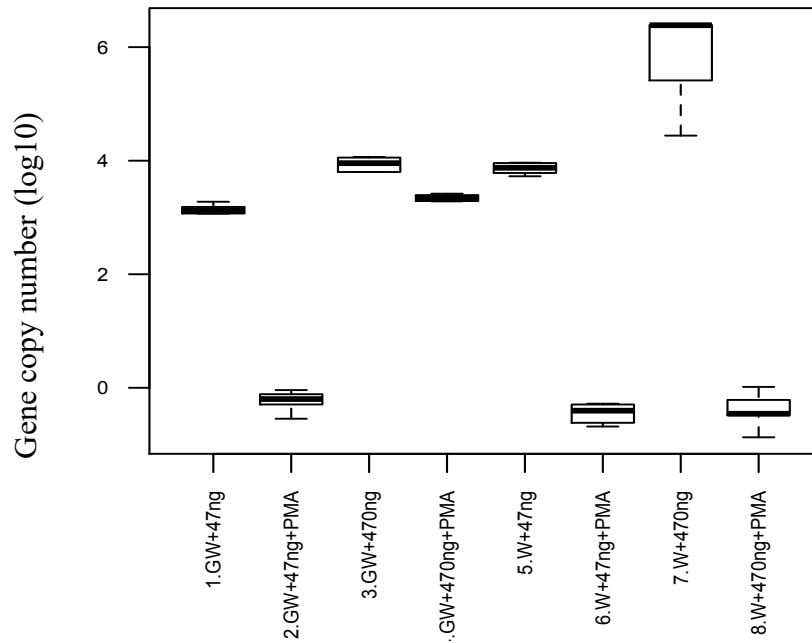


Fig. 5.3 Results of initial PMA tests with *E. coli* cultures. Serially diluted cultures ranging from 10⁵ to 10⁷ cfu/mL were either (LU) live and untreated, (HK) heat killed and untreated, (LP) live and PMA treated, or (HKP) heat killed and PMA treated. During gel loading, two samples were accidentally loaded in opposing gel lanes (heat killed, PMA treated sample (10⁵) was incorrectly loaded in the “untreated” gel, and vice.

A)



B)

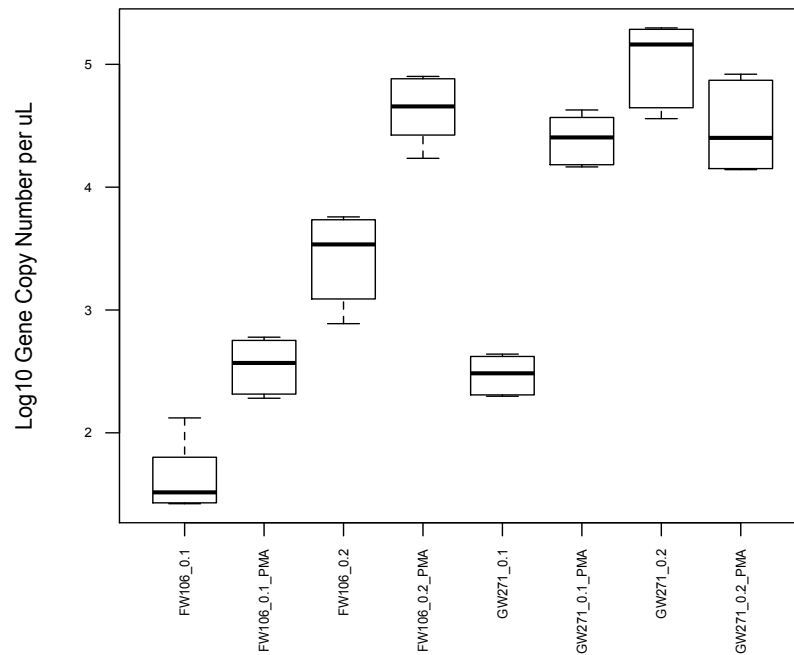


Fig. 5.4. A) qPCR results of spiked *E. coli* DNA experiments demonstrating that 10uM PMA removed 47ng of *E.coli* DNA spiked into representative groundwater. However, 10uM of PMA was not sufficient to effectively remove 470ng of *E. coli* DNA spiked into representative groundwater, although it was sufficient to remove the same amount from

filter sterilized water. These results suggest that while there may be some interference in groundwater that may prevent PMA-DNA binding at higher DNA loads than would be expected for this study, 10uM is sufficient for non-contaminated used in this study groundwater used in this study. B) Boxplots representing the range of amplified 16S rRNA gene copies from DNA extracts of PMA untreated and treated groundwater filters.

Abundances of Dominant Populations Within Groundwater May Be Inflated by Non-Viable DNA While Rare Populations May be Underestimated

In the 0.1µm fraction of uncontaminated groundwater GW271, there was a reduction in the relative proportions of two dominant OTUs at the family level most closely related to family *Egigoccaceae* (Phylum Actinobacteria) (Fig 5.4). In non-PMA treated samples of GW271, the cumulative representation of these two OTUs summed to roughly 60% of the total community. However, for PMA treated samples the cumulative relative abundance of *Egigoccaceae* decreased by nearly half (Fig 5.4). Coinciding with reduced relative abundances of *Egigoccaceae*, two OTUs increased in representation. An OTU most closely related to family *Acidiferrobacteraceae* increased in observed relative abundance, from roughly 4% in untreated samples to 20% in PMA treated samples. Additionally, an OTU most closely related to order *Desulfobacteraceae* increased from less than 0.1% to nearly 6% relative abundance.

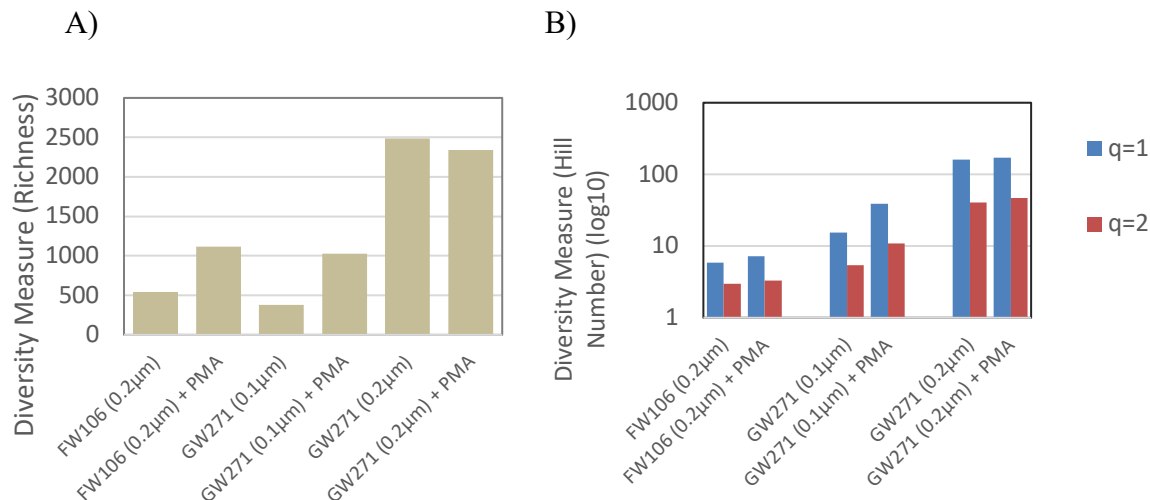


Figure 5.5. A) Richness and B) diversity measures based on Hill Numbers for PMA treated and untreated samples.

SIMPER analysis was used to calculate the Bray-Curtis dissimilarities between individual populations of treated and non-treated samples (Fig 5.5). Based on SIMPER results, differences between abundances of Actinobacteria populations contributed most to dissimilarities between treated and non-treated samples for both 0.2µm and 0.1µm fractions of non-contaminated groundwater (GW271). Overall, Actinobacteria were observed less abundant in PMA treated samples. This indicates DNA from these Actinobacteria OTUs may have been partially derived from non-viable sub-populations and/or relic DNA. The removal of non-viable DNA from Actinobacterial populations resulted in increased representation of other populations (*e.g. Acidiferrobacteraceae*, and *Desulfobacteraceae*) (Fig. 5.4 and 5.5).

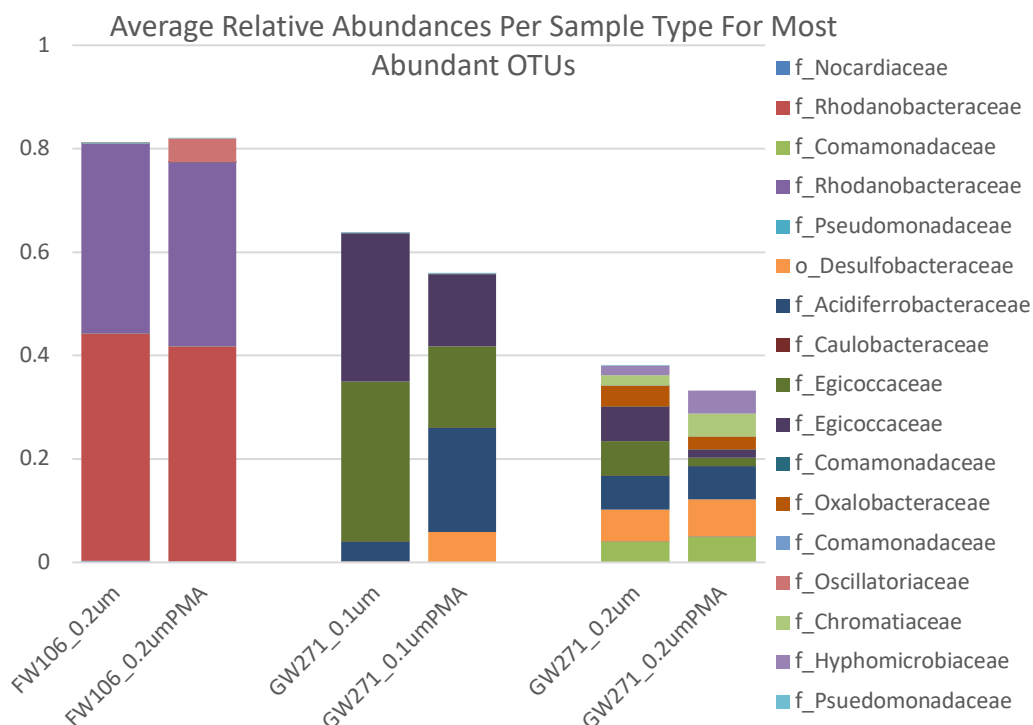


Figure 5.6. Bar plot of comparisons between average relative abundances of PMA treated and non-treated samples for 0.2µm and 0.1µm filtered fractions of sampled contaminated (FW106) and uncontaminated (GW271). Filters of the 0.1µm size fraction are not shown as there were insufficient DNA yields or sequence data for proper analysis.

In contaminated groundwater (FW106, 0.2µm fraction), the most abundant populations were two OTUs classified as family *Rhodanobacteraceae*. However, in PMA treated samples, an additional OTU of *Oscillatoriaceae* increased in representation within the sampled community of the 0.2µm filter fraction. Even still, the representation of *Rhodanobacteraceae* did not decrease considerably, despite SIMPER analysis indicating a high Bray-Curtis dissimilarity between treated and non-treated samples (Fig. 5.6). As SIMPER analysis calculates the Bray-Curtis dissimilarities of total counts (not relative proportions), high dissimilarity values between PMA treated and untreated samples may be a reflection of differences in individual counts and total library sizes (data not shown).

Influence of Non-Viable DNA on Bacterial Abundance Estimates in Sediments May be Depth Dependent in an Uncontaminated Core

The microbial diversity from freshly cored uncontaminated sediment was compared across three depths representing three distinct hydrological zones (unsaturated, saturated, and transition) of freshly cored sediment from a noncontaminated area were compared. The intermediate depth had the highest total cell densities (Appendix D, Supplementary Figure D.3) and was the most diverse based on SSU rRNA gene abundances (Fig. 5.7). In contrast, deeper sediments from the saturated zone displayed lower diversity and were dominated by few populations. In both unsaturated and transition zone sediments, PMA treatment resulted in a decrease of the most abundant populations (*Enterobacteraceae* and *Ktedonobacteraceae*, respectively). In contrast, the most abundant populations within saturated sediments (*Pseudomonadaceae*, *Enterobacteraceae*, and *Commamonadaceae*) remained dominant after PMA treatment.

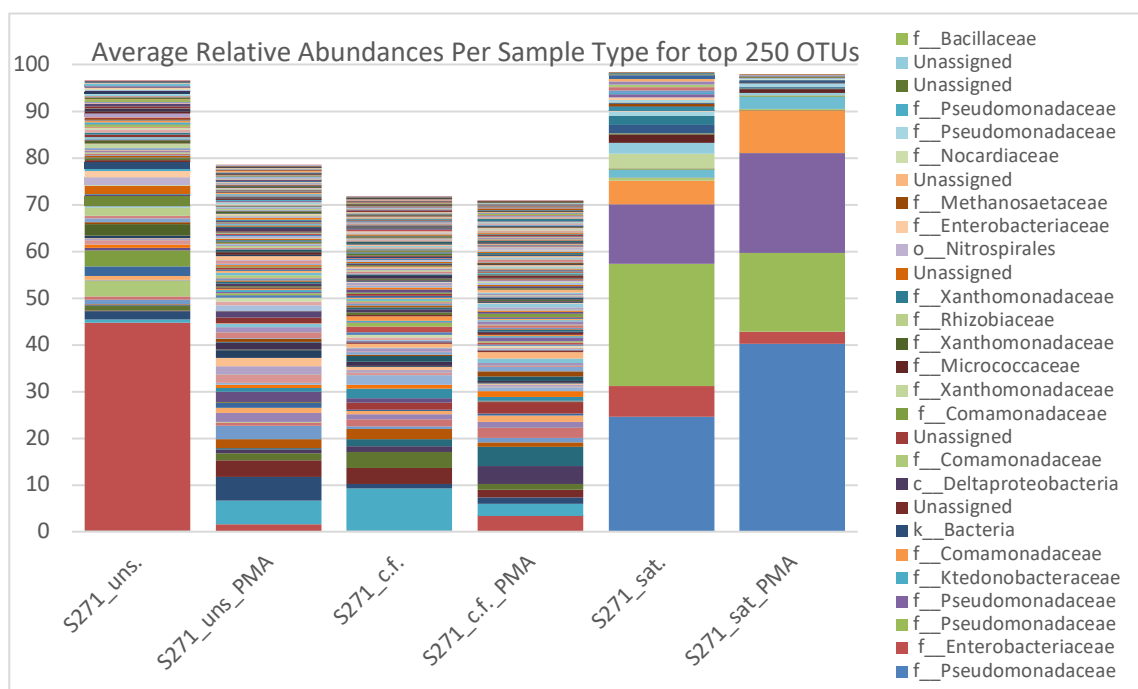


Figure 5.7. Bar plots of relative abundances of OTUs in PMA treated and untreated cored sediment representing three vertical depths (uns.= unsaturated, c.f.= capillary fringe sat.= saturated) for uncontaminated core EB271 (S271).

Discussion

Culture independent DNA based methods have allowed microbial ecologists to survey the vast diversity that exists in natural systems. However, a limitation of DNA based methods is that extracted nucleic acids can originate from cells at various physiological states. This limitation may result in misleading assessments of diversity, resiliency, and subsequent inferences regarding ecosystem functioning at a given point in time (Jones and Lennon, 2010; Carini et al., 2016). DNA from dead or lysed cells has been shown to persist in natural systems (Nielsen, 2007), and depending on the environment, cells can remain dormant for many years (Jones and Lennon, 2010). Techniques capable of distinguishing between different pools of DNA within environmental samples can be beneficial in microbial ecology, particularly for environments that may show substantial differences between viability of resident microorganisms, and thereby provide a better assessment of the “true” genetic capacity of the system.

Recently it was shown that relic DNA has the potential to inflate microbial diversity estimates for certain soil types (Carini et al., 2016). However, in lake ice, deep seafloor sediment, and municipal wastewater, the contribution of extracellular DNA to diversity measurements was observed to be negligible (Knocker et al., 2007; Dieser et al., 2010; Ramirez et al. 2018). In this study, both uncontaminated and contaminated groundwater showed insignificant differences between the richness (Fig. 5.3C) and minimal differences

in diversity (Fig 5.3D) between PMA treated and untreated samples. With few exceptions, dominant populations remained abundant regardless of PMA treatment, indicating that abundant groundwater populations of a shallow, terrestrial aquifers were mostly viable or at least “not dead”. However, the presence of rare members within groundwater communities within some non-PMA treated samples increased in proportional representation after PMA treatment (Fig 5.4, 5.5), indicating that there may have been enough removal of DNA from non-viable members of abundant communities that allowed for the detection of less abundant community members.

Interestingly, contaminated groundwater displayed significant differences in detectable gene copy number in samples treated with PMA (Fig. 5.4, Appendix D, Supplemental Figure D.1.B). Counterintuitively, the PMA treated samples in this study had significantly higher amplifiable gene copy numbers than untreated samples. This observation has been reported previously in acidic and clay rich soils (Carini et al., 2016). It is possible that the environmental properties of contaminated groundwater, including pH, may affect properties of the PMA molecule (Andreas Nocker, personal communication). For example, fluorescent signal has been shown to be highly pH dependent for several fluorescent dyes (Baldock et al., 2013). The contaminated groundwater used in this study was acidic (pH=3) and visibly turbid as compared to uncontaminated groundwater. Although results showed that there were no significant effects on richness or diversity between treated and non-treated samples from contaminated groundwater despite significantly higher SSU rRNA gene copy number in contaminated samples, the effects of

contamination and pH on PMA behavior (*e.g.* permeability, DNA binding, photo-reactivity) were not directly tested and still need to be fully elucidated.

While sediments of the shallow subsurface have been documented to contain a rich diversity of microbial organisms (Hug et al., 2016, Anantharam et al., 2016), a general trend has been observed of decreased cellular abundances and diversity in relation to vertical depth. It has also been shown that cellular counts, richness, and diversity can increase in localized “hotspots” (Smith et al., 2018, references therein). However, the extent of moribundity and death at various depths of the subsurface profile is not currently known. In this study, distinct community profiles were observed based on depth in a non-contaminated sediment core representing three distinct hydrologic profiles. Sediments from the intermediate transition zone were observed to have greater overall richness and diversity than either unsaturated (*i.e.* vadose) and saturated sediments, and showed minimal differences between PMA treated and untreated samples. Overall, sediments from the saturated zone had fewer total sequences, richness, and diversity than those of shallower depths, corroborating the previously observed general trend of decreasing abundance and diversity over depth. However, shallower depths may have had a greater proportion of non-viable DNA than those from deeper depths. The shallowest depth analyzed (137.16 cm-bgs, representing the unsaturated vadose zone) showed the greatest shift in community composition between PMA treated and untreated samples, which resulted in a large reduction in a highly dominant population of *Enterobacteraceae*. While it is possible that insufficient amounts of PMA were used in sediment samples to fully distinguish between possibly viable and non-viable populations, the current results suggest that most

populations across all depths are mostly viable, at least for non-contaminated sediments. It is currently unknown if the same trend is true for contaminated sediments. Contaminated sediments are currently being processed and analyzed in order to assess richness and diversity over depth and possible influences of non-viable DNA on these estimates.

Overall, the results of the above study demonstrate that the majority of bacterial cells sampled from both contaminated and uncontaminated subsurface groundwater and sediments were mostly intact. These results are useful for future studies that will aim to differentiate between active states and activity rates of bacterial cells across spatial and temporal scales and degrees of environmental stresses such as contamination.

EPILOGUE

Global demands for natural subsurface resources continues to grow at increasing rates. However, we currently lack a thorough understanding of how extraction and utilization of underground resources [fertile soils (Tilman et al., 2011), clean and potable groundwater (Rulli et al., 2013), minerals (Kesler, 2007), and fossil fuels (Jackson and Smith, 2014)], impact terrestrial subsurface ecosystems. Subsurface microorganisms have been linked to critical processes such as the formation of soils (Lynch et al., 1985), maintenance of potable groundwater (Chapelle, 1993), and the production of geothermal energy (Gniese et al., 2014). Despite the importance of subsurface bio-systems, we know relatively little about the vast and diverse ecological network of subsurface microorganisms or the environmental parameters that constrain their activities and distributions. In order to improve prediction and management strategies for subsurface environments, as well as to accurately assess the extent of natural temporal variation that should be expected in healthy systems, a thorough understanding of the structure, function, and environmental drivers of native microbial communities over space and time is critical. The work presented in this dissertation aimed towards providing initial assessments of the variability of microbial diversity over temporal (Chapter Two) and spatial (Chapters Three, Four, Five) scales in a shallow aquifer.

Historically, ecological surveys of microorganisms have been performed via one-time observations, providing a “snap shot” of richness and diversity at a single point in space and time (Gonzalez et al., 2012). Environmental collection of data can require large investments in resources and funding, especially for longer-term studies (Field et al., 2007;

Hughes et al., 2017). Identifying the field-relevant time-scales with which an ecological process of interest should be studied (*e.g.* hours, days, months, years), especially for microorganisms, can also be challenging (Nemergut et al., 2013). Irregularly sampled environments due to a variety of logistical difficulties with sampling efforts lead to missing data, resulting in datasets that must be manipulated statistically (*e.g.* interpolation) for many time-series analyses to perform well (Rehfeld et al., 2011; Legendre and Legendre, 2012). Although technological advances in molecular-based technologies have facilitated investigations at greater temporal and spatial resolutions for microorganisms (Fierer, 2008; Gonzalez et al., 2012 and references therein), additional complexities with respect to the computational and statistical tools required to analyze spatiotemporal data have challenged ecologists (further discussed below).

Despite challenges associated with temporally collected data, the concern over altered ecosystems on a global scale due to anthropogenic disturbances, and their ability to continue to sustain life has pressed ecologists to understand biodiversity in relation to time (Wolkovich, 2014). Within the last 10 years, spatiotemporal studies of microbial diversity have provided greater insight into short- and long-term variability of microorganisms *in-situ*. Dissimilarity metrics have been used to show that conditionally rare taxa accounted for the majority of temporal variation in microbial beta-diversity in a variety of environments (Shade et al., 2014). Other studies have identified patterns related to seasonalities in marine systems (Gilbert et al., 2014), and human gut microbiome succession (Koenig et al., 2011). Despite the gains in the understanding of spatiotemporal microbial biodiversity, most studies relating to temporal dynamics have been performed

on surface systems, while temporal studies in less accessible subsurface environments are not as common.

In Chapter Two of this dissertation, SSU rRNA gene paired-end sequencing was used to observe groundwater community structural shifts at varying levels of intensity over time in a non-contaminated area of the EOR-FRC. While natural variation in groundwater is likely and to be expected (Nelson, 2002), it was originally assumed that observed shifts in community structure would be generally similar across groundwater collected from the three sampled field wells as they i) are in relatively close proximity (between 10 and 20 m apart), ii) tap into the same uncontaminated aquifer, and iii) have the same underlying geology. Surprisingly, significant differences over time were observed in the populations present, the proportional representation, and associations between geochemical and population variation, both within and across field wells.

It is generally assumed that greater diversity has a direct and positive effect on ecosystem functioning (Reese and Dunn, 2018); however, conflicting results have been observed when using experimental approaches on natural systems and *in-vitro* microcosms (Griffiths et al, 2001; Franklin et al., 2001; Bell et al., 2005). In an uncontaminated aquifer, three different sampling locations showed variable diversity over short-time scales (~3 d). Potential functionality (as determined by functional gene array), however, did not show similar variations over time. While correlation- and regression analysis (*i.e.* SPARCC, linear-mixed effects models) and constrained ordinations (*i.e.* DCA) were utilized in order to infer species-species and species-environment relationships, the increased complexity

of sequentially collected data required the use of statistical methods that were more suitable for the unique challenges related to temporal analysis.

Unlike experimentally controlled studies, a series of data collected sequentially over time in natural systems may have inherent internal structure, such as trends, seasonalities, non-stationarity, and autocorrelation that must be identified and accounted for in order to avoid spurious or misleading inferences regarding species-species or species-environment relationships (Legendre and Legendre, 2012). For example, autocorrelation, or serial-dependence, is an inherent component of data collected over time (Lutkepohl, 2005). When autocorrelation is strong, many traditional statistical methods that rely on independent observations are rendered invalid. It has been argued that ecologists would benefit from borrowing statistical models that have originated from scientific disciplines outside of ecology in order to meet the challenges of analyzing complex spatiotemporal data at both short- and long-term scales (Anderson et al., 2008; Wolkovich et al., 2014).

Largely due to their simple structure, flexibility, and applicability, Vector autoregression (VAR) models are one of the most popular and widely used models in econometrics (Lutkepohl, 2005). VAR models are commonly used to describe and predict dynamic behavior of economic processes, such as fluctuating stock markets (Papapetrou, 2001), and local prices of commodities in relation to global demands (Nakajima, 2014). VAR models have also been adopted by other scientific disciplines to study temporal dynamics, including hydrogeology (Mangiarotti et al., 2012), human physiology (Matsukawa and Wada, 1997), and marine microbial ecology (Lui et al., 2007). VAR

models allow for the identification of and adjustment for inherent structure within temporal data sets (*i.e.* trends, non-stationarity, autocorrelation). When paired with null-hypothesis testing (*i.e.* Granger Causality), significance of associations and their potential directionality of relationships (*e.g.* “causal” changes in the first series that precede and helps to better predict changes in a second series) can be inferred (Faust et al., 2015). The ability to detect potential time-delayed relationships is also a useful feature of VAR models.

In an attempt to infer diversity-function relationships and directionality between geochemistry and fluctuating diversity in this study, Vector Autoregression and Granger Causality were used to associate changes to bacterial diversity (as measured by taxa richness) and changes in measured geochemistry. In most cases, relationships between geochemistry and richness were weak except for communities sampled from a newly cored field well (FW305). However, as the observation period was relatively short (12 weeks), and confined to the transition period between two seasons (summer-fall), it remains unclear how these short-term fluctuations relate to the long-term stability (or instability) of the aquifer. Additional short-term studies that are equally spaced to identify potential effects of natural diel cycles on water table fluctuations (*e.g.* increased evapotranspiration by surface vegetation at hotter periods of the day), combined with longer-term studies that can identify seasonality effects, are worth further investigation.

It has recently been demonstrated that inferences made about subsurface community dynamics based solely on the planktonic fraction may not be adequate or fully representative of subsurface bacterial diversity (Hug et al., 2015). Therefore, in order to

fully investigate the diversity within subsurface environments, it is essential to study both planktonic and attached populations. Towards this aim, Chapter Three presents results from surrogate sediment samples that were incubated in field wells (concurrent with the temporal groundwater study from Chapter Two) in order to assess the sediment-associated bacterial fraction of the groundwater environment. The sediment-associated bacterial populations observed were comprised of a subset of populations observed in groundwater, which corroborates previous findings that examined both attached and planktonic fractions (Hazen et al., 1991; Reardon et al., 2004; Kellerman and Griebl, 2012). With little exception, the community structure observed on within-well Biotrap® samplers were highly similar despite a high degree of temporal fluctuation of source populations in the groundwater. The reproducible observation suggests a strong selection pressure towards sediment-associated/attached lifestyle from a subset of groundwater populations. However, it is possible that the reproducibility of surrogate sediment samplers may have been a consequence of i) the length of incubation time (three-months), ii) the period of observation under study (*e.g.* drier months of late summer/early fall), and/or iii) the homogenization of bulk surrogate sediment within each sampler prior to nucleic acid extraction. Although all samplers were cut in half and each respective half provided the same results. Nevertheless, the work presented here suggests that after three months of incubation over a relatively dry period (July-October), separate surrogate samplers are mostly reproducible when processed via pulverization/homogenization prior to DNA extraction. Future studies could be designed to take advantage of this reproducibility. For example, samplers could be

sequentially harvested over an equidistant interval (*e.g.* weeks) to investigate colonization and succession events as they relate to source microbial fluctuations.

The sediment-associated fraction of subsurface communities have traditionally been investigated using bulk analyses such as those described above. However, bulk analyses have the potential to obscure the particle-scale heterogeneities that can directly influence the micro-environments in which microorganisms reside. In Chapter Four, the possible influence of particle-size as a driver of community structure and assembly was investigated. Results from these studies suggest that overall, particle-size might influence community richness and diversity in an environment-dependent manner. Communities associated to particles incubated in three different environment types (laboratory reactors, a field well previously characterized as highly variable in terms of source populations and geochemistry, and a second field well previously characterized as more relatively stable) were influenced differently by particle-size effects. Communities observed on large particles sizes (3000 μm) from more relatively 'homogenous' environments (*e.g.*, laboratory reactors and stable field well) were comprised of more total rare populations (relative abundances $<0.01\%$). It is possible that the different hydrogeological and microbial dynamics in temporally homogenous versus heterogeneous environments results in unique behavior in terms of source populations, biofilm formation, and subsequent porosity, permeability and cellular dispersal at the particle scale. For example, larger beads might have different average pore velocities compared to smaller bead sizes that allows for a wider array of different populations (*e.g.*, cells that grow more slowly). Further work to

experimentally address the effects of hydrological flow, porosity, permeability, and source populations could provide greater insights into these dynamics.

Results of molecular survey methods that depend on nucleic acids have the potential to be misleading due to the presence of extracellular or relic DNA that is co-extracted from environmental samples. Chapter Five attempted to address this concern and determine the potential impact of extracellular, or relic DNA on diversity estimates of subsurface groundwater and cored sediments from the tested field samples. Results indicated that the influence of extracellular DNA is mostly minimal, although rare populations may have the potential to be underestimated. As rare population members have been linked to biogeochemical cycles (Jousset et al., 2017 and references therein), the understanding of their presence and distribution across environments can be aided by culture-independent methods that can increase detection, such as PMA-Seq.

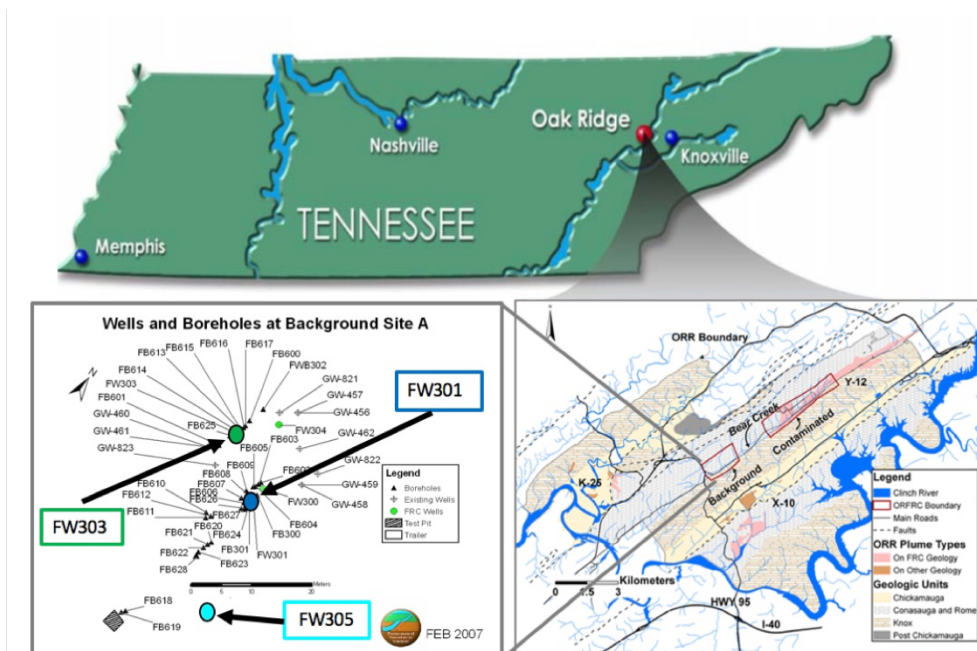
In conclusion, the results presented in this dissertation contribute to the collective understanding of subsurface bacterial diversity of shallow terrestrial aquifers. Additionally, the presented results have field-relevant implications for the sampling, estimation, and expected variability of microbial populations within uncontaminated shallow aquifers.

APPENDICES

APPENDIX A

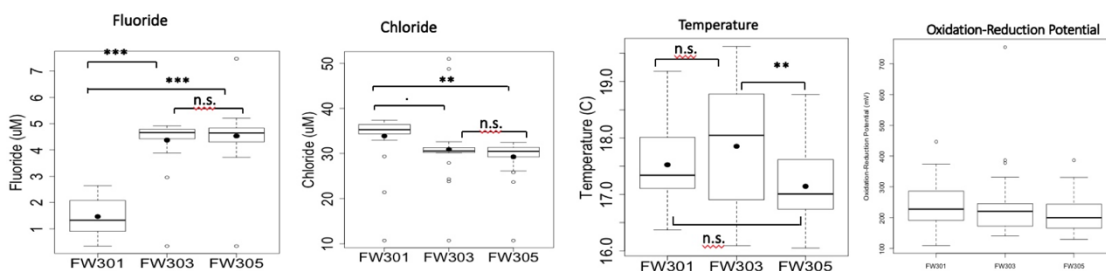
SUPPLEMENTARY MATERIAL FOR CHAPTER TWO

Supplementary Figure A.1.



Geographic location of non-contaminated background site of the EOR-FRC and chosen field wells. All field wells have the same underlying geology, and are located between 10 – 25m apart.

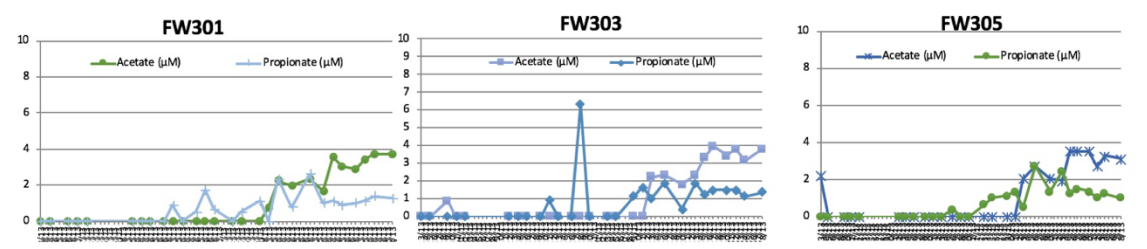
Supplementary Figure A.2. (a-d)



Tukey-style boxplots of within-well geochemistry for three wells with $n=28$ for each well. Mean and median are denoted by a filled-in dark circle and horizontal line, respectively, within each boxplot. Significance testing was carried out via a one-way ANOVA. Wells

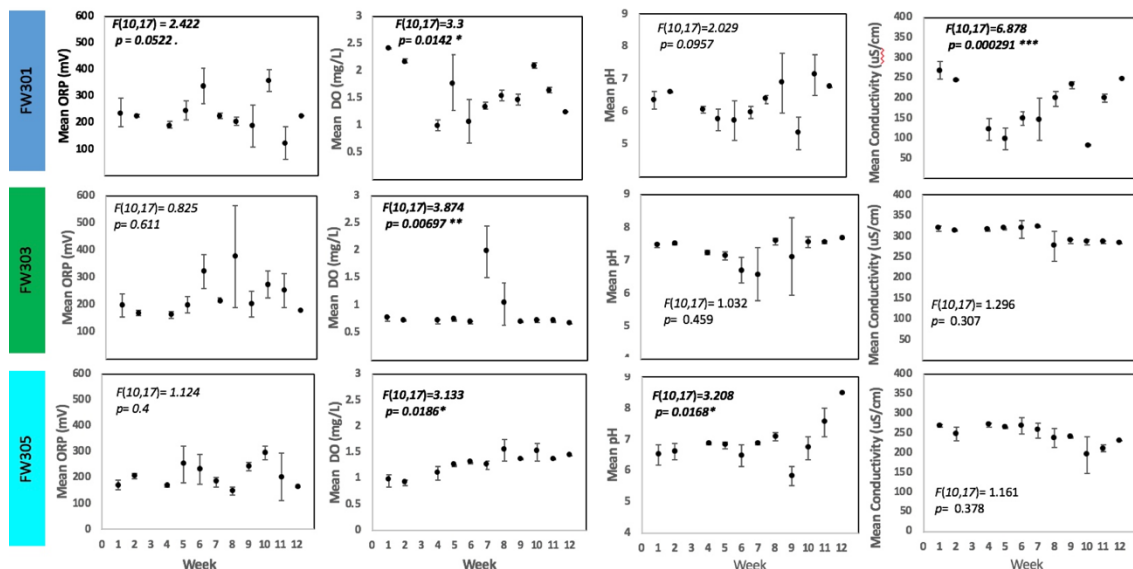
differed significantly in Fluoride ($F(2,81) = 106.6$, $p\text{-value} = 2\text{e-}16$), chloride ($F(2,81) = 151.19$, $p\text{-value} = 0.01$), and temperature ($F(2,81) = 4.785$, $p\text{-value} = 0.0109$), but not for ORP. Post-hoc analysis (pairwise t-tests with holm adjustment) were performed to differentiate differences between groups. FW301 was significantly different from both wells in Fluoride and Chloride levels, while FW303 and FW305 were not significantly different from each other. Temperature in FW305 was significantly different from FW303, but not FW031. Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1.

Supplementary Figure A.2 (e-g)



Scatter plots of measured organic acid concentrations over time in sampled groundwater from each well

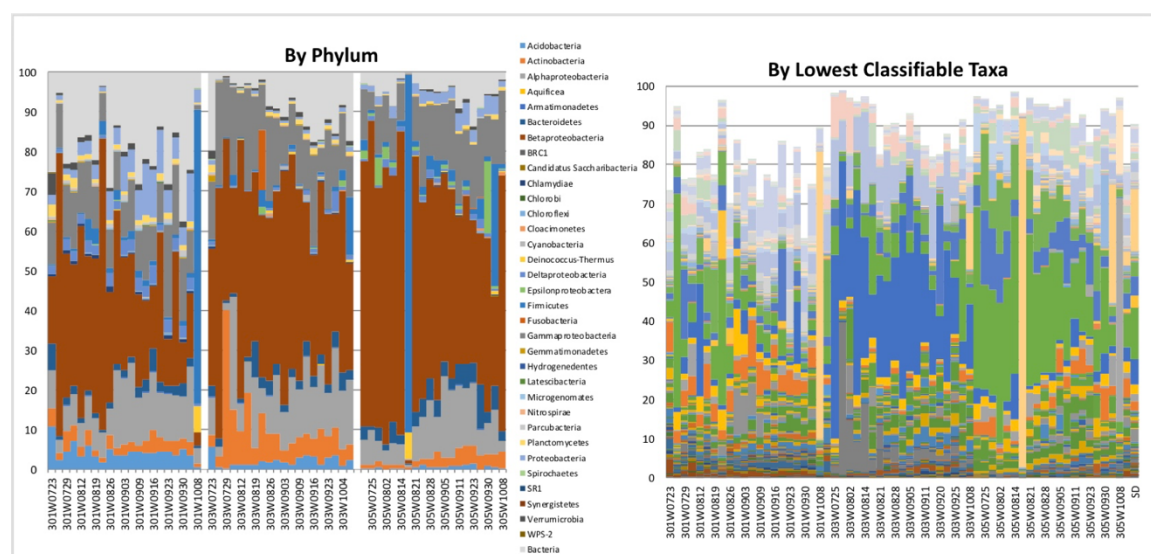
Supplementary Figure A.3.



Plots show weekly means and standard errors (SEM) within-well groundwater geochemistry over 11 weeks for Oxidation Reduction Potential, Conductivity, pH, and Dissolved Oxygen. Sample sizes per well are $n=2$ for weeks 1, 7, and 9, and $n=3$ for all other weeks except week 12, in which only 1 sampling event occurred ($n=1$). Week 3 has

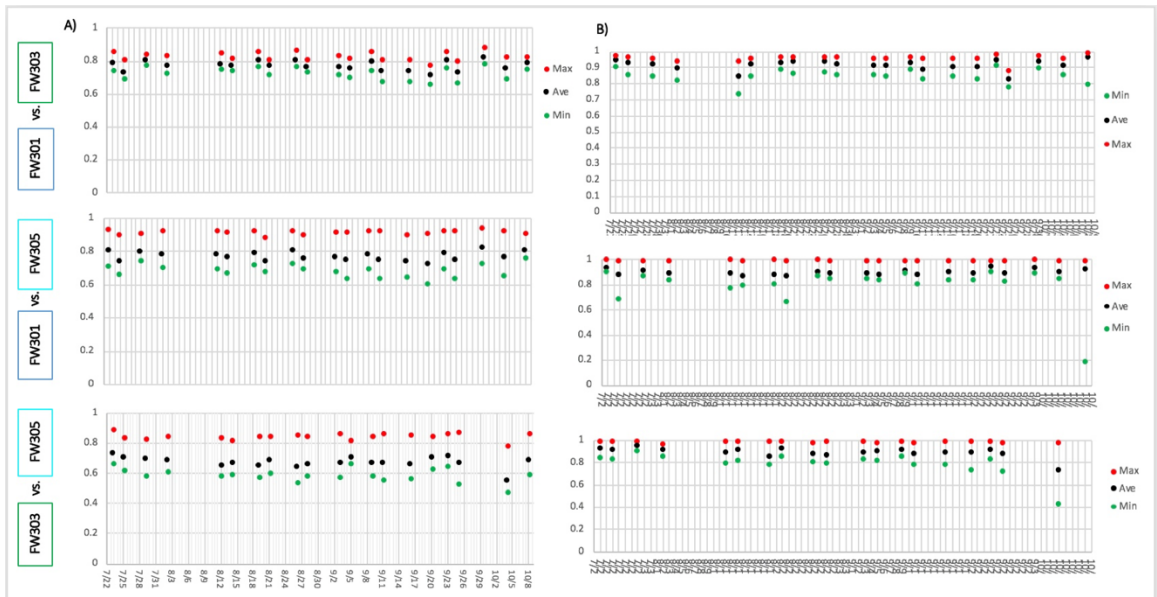
no plotted data because no samples were collected that week. Test of significance was carried out via a one-way ANOVA in R and a Tukey's HSD test on all samples taken over 11 weeks for a total of 28 samples per well (N=28). FW301 shows significantly different weekly means for measured ORP ($F(10,17) = 2.422$, $p\text{-value} = 0.0522$), DO ($F(10,17) = 3.33$, $p\text{-value} = 0.0142$), and Conductivity ($F(10,17) = 6.878$, $p\text{-value} = 0.000291$). FW303 and FW305 both varied significantly in measured dissolved oxygen ($F(10,17) = 3.874$, $p\text{-value} = 0.00697$; and $F(10,17) = 3.133$, $p\text{-value} = 0.0186$, respectively). Additionally, FW305 varied in measured pH ($F(10,17) = 3.20$, $p\text{-value} = 0.0168$). Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '.' 1.

Supplementary Figure A.4.



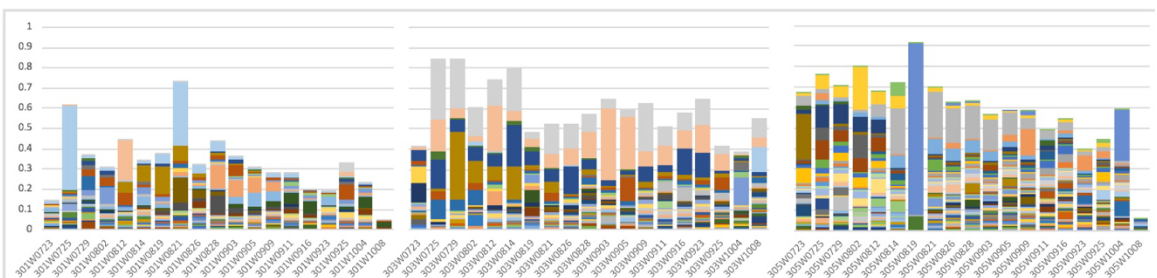
Relative abundances by A) Phylum and B) lowest classifiable taxa. Legend corresponds to Phylum bar chart. Phylum Proteobacteria is further broken down into classes.

Supplementary Figure A5.



Minimum, average, and maximum dissimilarities of sampled groundwater OTU diversity across wells. A) Sorensen dissimilarities shows that on average, wells are approximately 80% dissimilar to each other based on OTUs present or absent over time. B) Bray-Curtis dissimilarities shows that on average, wells are between 68-96% dissimilar in the OTUs abundances of present OTUs from day to day. Sparse OTUs (not present in >50% of all groundwater samples) were removed prior to Bray-Curtis analysis.

Supplementary Figure A.6.



Relative abundances per sample for the OTUs that contributed $\geq 1\%$ in at least one paired-week Bray-Curtis dissimilarity based on SIMPER analysis for rarefied datasets. A total of 55 unique pair-wise dissimilarities pairings were calculated. A) In FW301, the cumulative abundances of OTUs is typically less than 40% of the total measured abundance for any given date. This indicates that the majority of OTUs contributing to weekly dissimilarities are moderately to rarely abundant on any given day, with period

blooms of different OTUs contributing largely to weekly dissimilarities. B) Unlike in FW301, the OTUs with the most contributions to weekly dissimilarities are often highly abundant in FW303. This indicates that differences in this well are attributed to changes in consistently abundant, and not rarely transient, populations. C) Similarly to FW303, changes in abundant OTUs contributed to weekly dissimilarities in FW305.

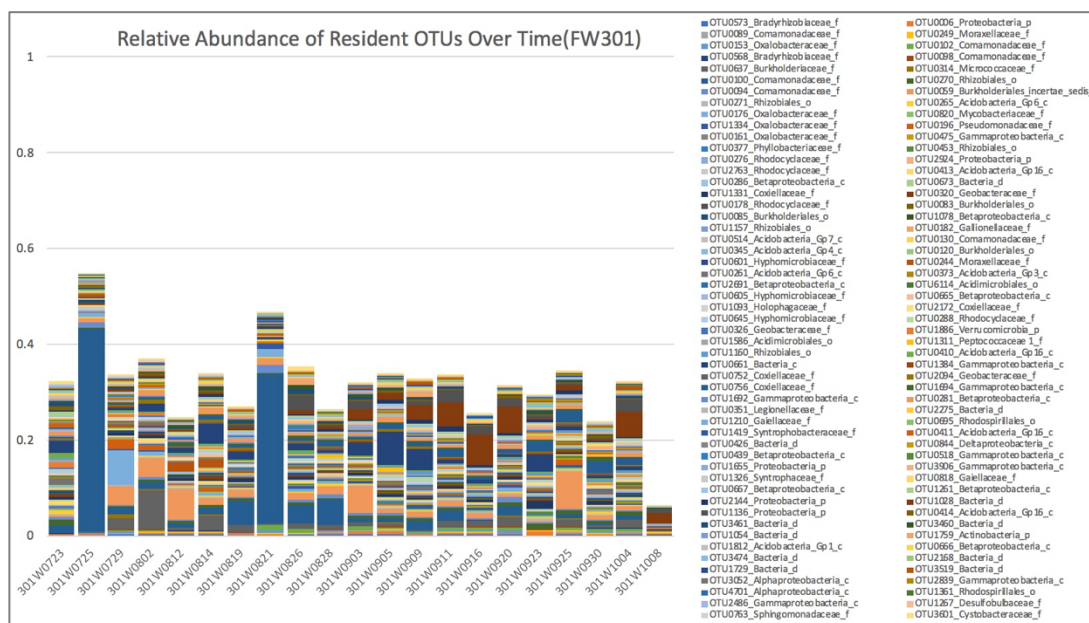
Supplementary Table A1. OTUs that contributed to ≥ 30 paired-week contributions in SIMPER analysis.

Taxon	Paired Week Contributions	Ave % Contribution	SD±	Taxon	Paired Week Contributions	Ave % Contribution	SD±	Taxon	Paired Week Contributions	Ave % Contribution	SD±
OTU0661_Betaproteobacteria_c	46	2.70	1.44	OTU0045_Burkholderiales_in_sed_f	55	10.20	3.74	OTU0089_Comamonadaceae_f	52	8.54	4.96
OTU1136_Proteobacteria_p	43	1.81	0.45	OTU0024_Burkholderiales_in_sed_f	54	10.45	7.87	OTU0290_Rhodocyclaceae_f	48	3.21	1.67
OTU0100_Comamonadaceae_f	42	10.31	8.58	OTU0719_Gammaproteobacteria_c	51	7.58	2.49	OTU0172_Oxalobacteraceae_f	41	4.65	3.56
OTU0059_Burkholderiales_in_sed_f	42	3.75	1.47	OTU0172_Oxalobacteraceae_f	46	6.21	4.67	OTU0100_Comamonadaceae_f	41	4.24	3.04
OTU1028_Proteobacteria_p	40	4.12	1.71	OTU0314_Micrococcaceae_f	43	7.92	7.94	OTU0176_Oxalobacteraceae_f	38	4.15	3.94
OTU0637_Burkholderiaceae_f	38	2.54	1.80	OTU0196_Pseudomonadaceae_f	36	3.24	1.83	OTU0322_Geobacteraceae_f	37	3.00	1.86
OTU2520_Rhodospirillaceae_f	37	5.55	3.59	OTU2023_Bacteria_d	34	4.04	1.82	OTU0094_Comamonadaceae_f	37	2.79	1.08
OTU0144_Burkholderiales_o	32	3.53	2.20	Total Contributing OTUs ($\geq 1\%$ Ave)	51			OTU0076_Comamonadaceae_f	34	3.93	1.37
OTU0290_Rhodocyclaceae_f	31	6.73	4.25					OTU0021_Burkholderiales_inc_se_f	33	2.17	0.72
Total Contributing OTUs ($\geq 1\%$ Ave)	51							OTU0637_Burkholderiaceae_f	32	1.78	0.55
								OTU0560_Caulobacteraceae_f	31	1.61	0.58
								Total Contributing OTUs ($\geq 1\%$ Ave)	72		

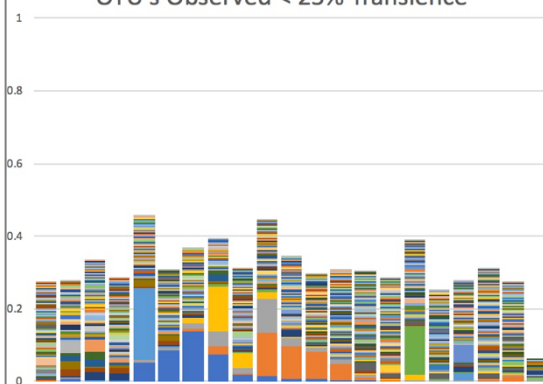
Supplementary Table A2. Estimated residence times based on hydraulic conductivity K of varying orders of magnitude using previous field data (Watson et al. 2004)

Well	Residence Time ($K = 1 \times 10^{-4} \text{ m/s}$)	Residence Time ($K = 1 \times 10^{-5} \text{ m/s}$)	Residence Time ($K = 1 \times 10^{-6} \text{ m/s}$)
FW301	3.9 hrs	39hrs (1.6 days)	390hrs (16.25 days)
FW303	3.9 hrs	39hrs (1.6 days)	390hrs (16.25 days)
FW305	7.9 hrs	79hrs (3.2 days)	790hrs (32.9 days)

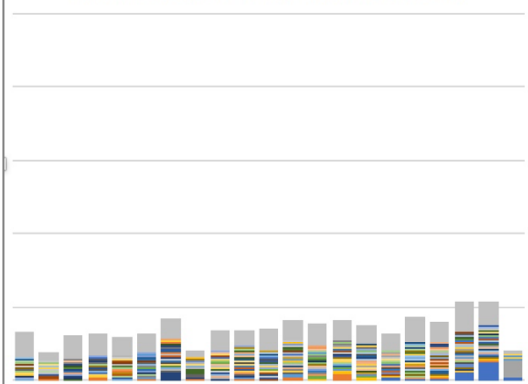
A)



B) Relative Abundance of Non-Resident OTU's Observed < 25% Transience

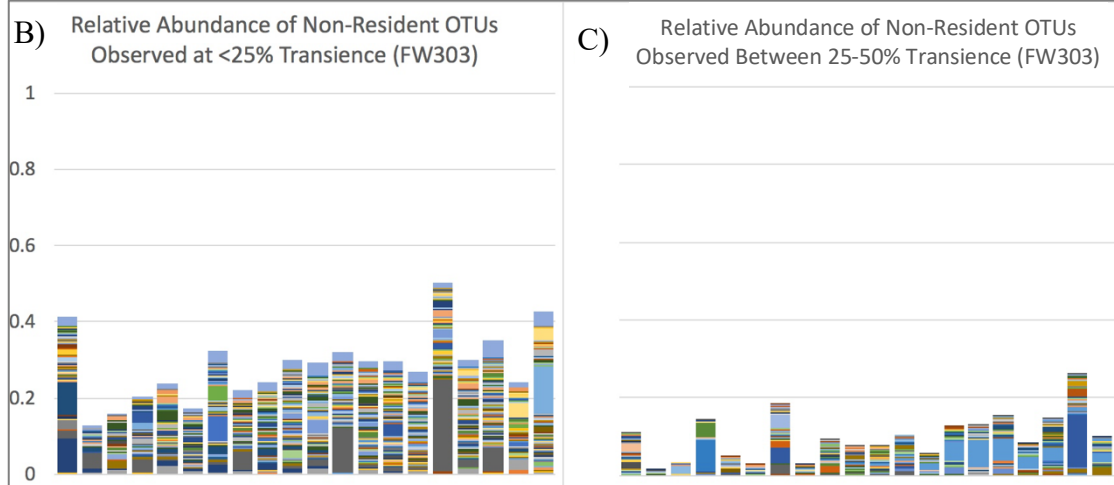
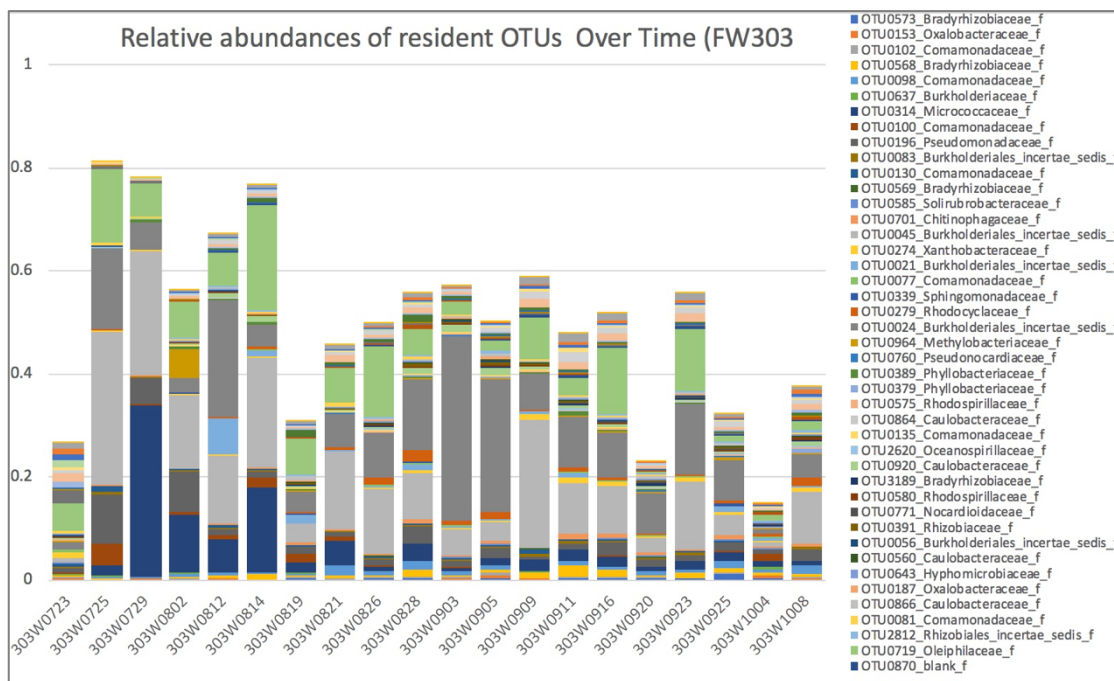


C) Relative Abundance of Non-Resident OTU's Observed Between 25-50% Transience (FW301)



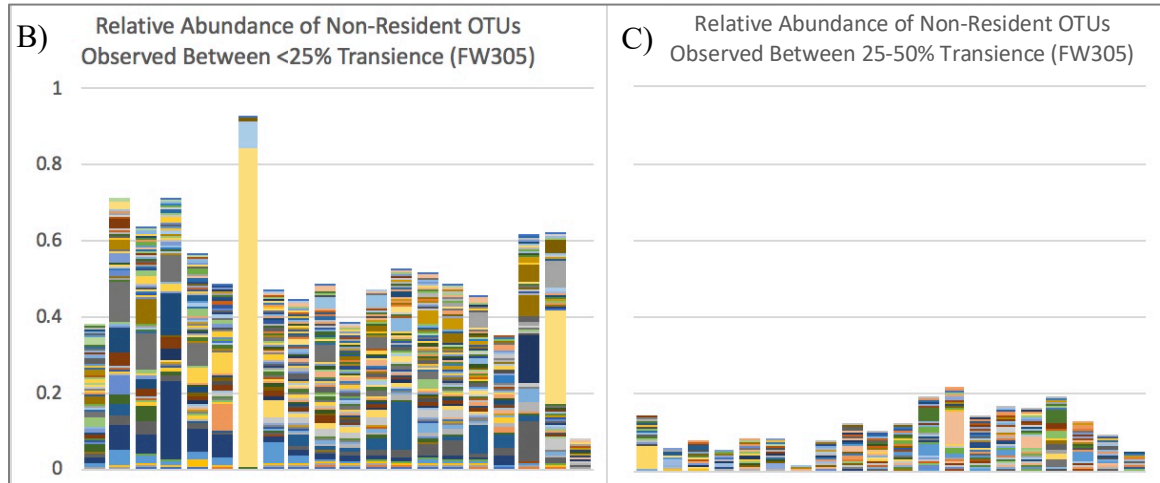
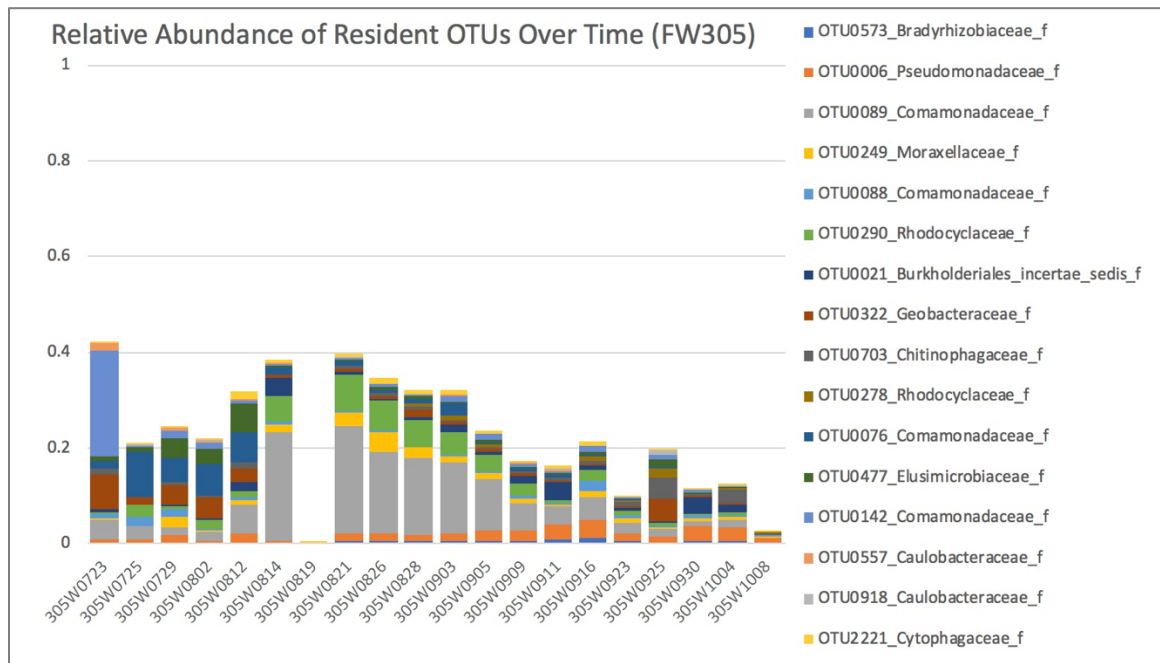
Supplementary Figure A7. Relative abundance bar charts for A) Resident and B,C) Transient OTUs observed in FW301 over time. FW301 had a total of 3,372 OTUs observed in that well over the three-month sampling period. A) OTUs were considered resident if they were observed in every sample analyzed for the three month sampling period. A total of 106 resident OTUs were observed over time in FW301. B) A total of 490 OTUs were observed in 75% of the dates sampled, and thus were considered transient in sampled FW301 groundwater at 25% transience. C) A total of 646 OTUs were observed in more than half of the dates samples, but less than 75% of the dates sampled, and therefore considered to be observed between 25-50% transience. The remaining 2,130 OTUs in FW301 were transient at greater than 50% (observed in less than half of all dates sampled).

A)

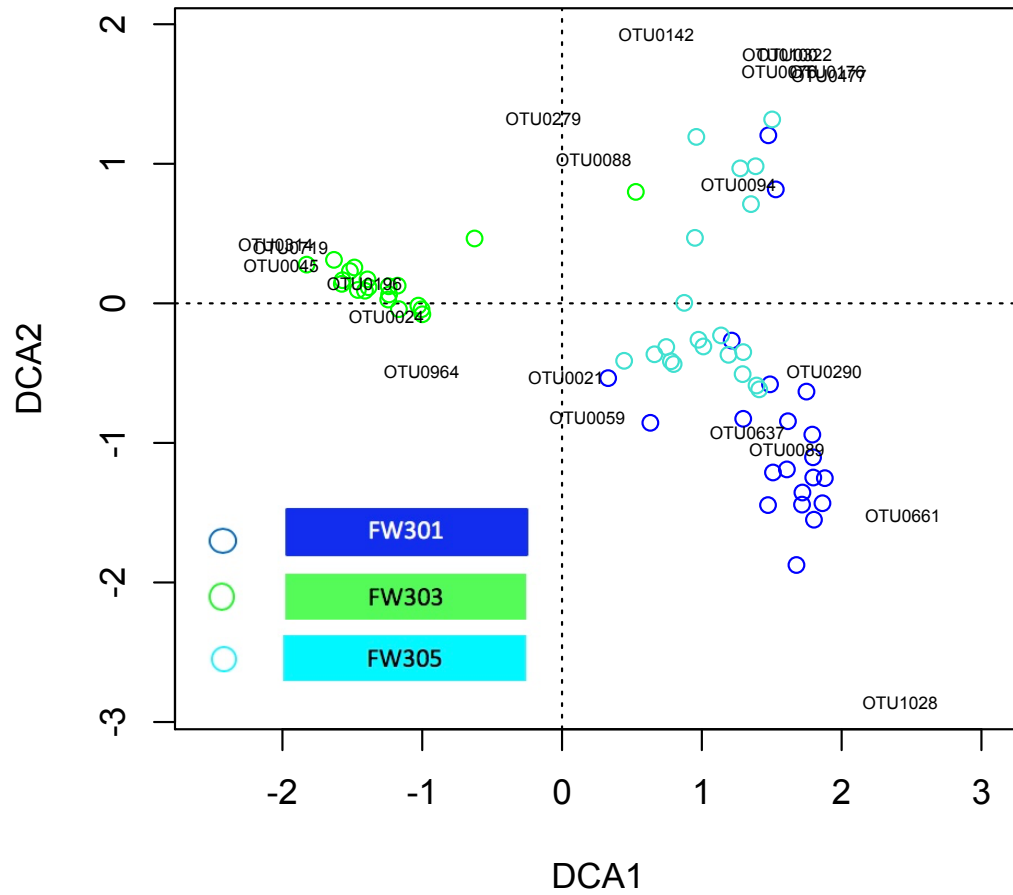


Supplementary Figure A8. Relative abundance bar charts for A) Resident and B,C) Transient OTUs observed in FW303 over time. FW303 had a total of 2,277 OTUs observed in that well over the three-month sampling period. A) OTUs were considered resident if they were observed in every sample analyzed for the three month sampling period. A total of 58 resident OTUs were observed over time in FW303. B) A total of 289 OTUs were observed in 75% of the dates sampled, and thus were considered transient in sampled FW303 groundwater at 25% transience. C) A total of 295 OTUs were observed in more than half of the dates samples, but less than 75% of the dates sampled, and therefore considered to be observed between 25-50% transience. The remaining 1,635 OTUs in FW303 were transient at greater than 50% (observed in less than half of all dates sampled).

A)



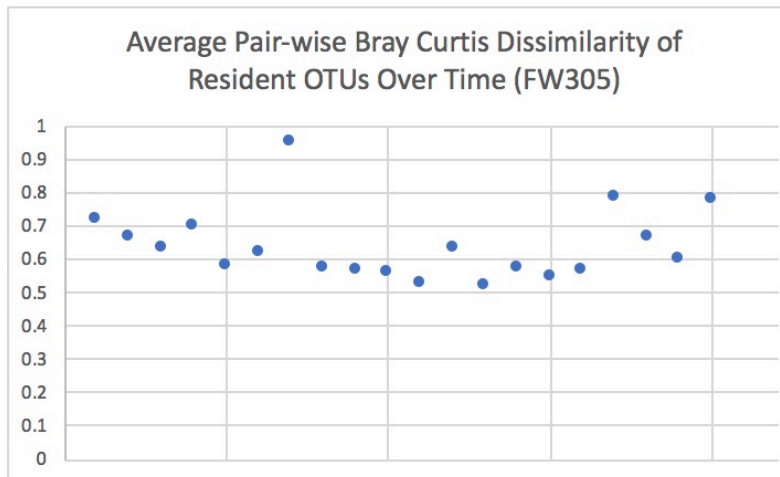
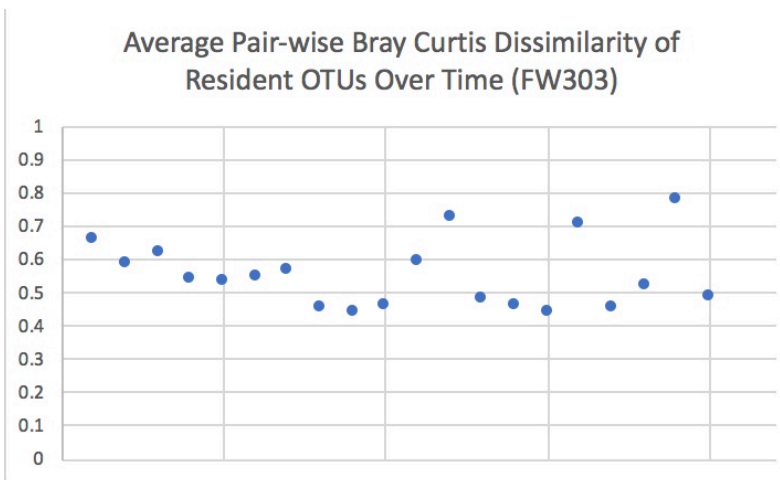
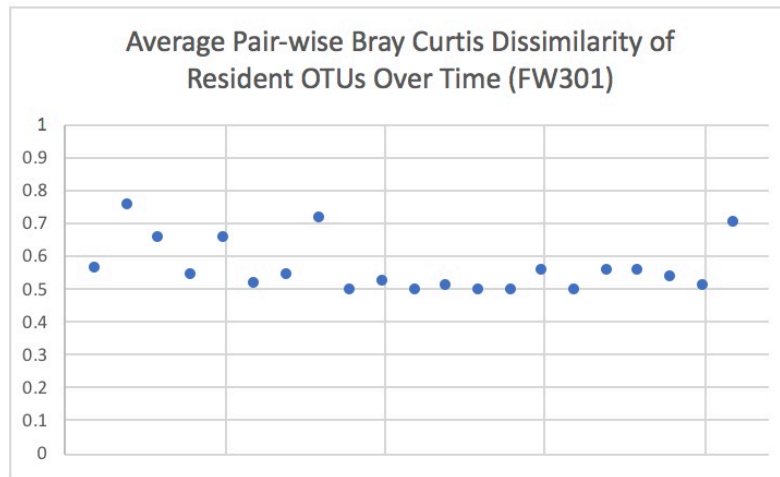
Supplementary Figure A9. Relative abundance bar charts for A) Resident and B,C) Transient OTUs observed in FW305 over time. FW305 had a total of 2,129 OTUs observed in that well over the three-month sampling period. A) OTUs were considered resident if they were observed in every sample analyzed for the three month sampling period. A total of 16 resident OTUs were observed over time in FW305. B) A total of 217 OTUs were observed in 75% of the dates sampled, and thus were considered transient in sampled FW305 groundwater at 25% transience. C) A total of 255 OTUs were observed in more than half of the dates samples, but less than 75% of the dates sampled, and therefore considered to be observed between 25-50% transience. The remaining 1,641 OTUs in FW305 were transient at greater than 50% (observed in less than half of all dates sampled).



Supplementary Figure A10. Detrended correspondence analysis (DCA) based on observed resident OTUs for each well. OTUs were considered resident if they were observed in every sample analyzed for the three month sampling period. FW301, FW303, and FW305 had a total of 106, 58, and 16 resident OTUs, respectively. Only the most abundant resident OTUs (greater than 5% relative abundance in at least one sample) are plotted. Wells are denoted by color (FW301, Blue; FW303, Green; FW305, Turquoise). The RDP taxonomic classifications for the most abundant OTUs plotted on the DCA are listed in Supplementary Table A3. Total Inertia of DCA plot = 3.682. Total percent variance explained by axis 1 and axis 2 is 20.5% and 15.4%, respectively.

Table A3. Most abundant (>5% relative abundance) resident OTUs per well and their taxonomic classification based on RDP alignment.

OTUId	Phylum	RDP C.I.	Class	RDP C.I.	Order	RDP C.I.	Family	RDP C.I.	Genus	RDP C.I.	Strongest DCA association:
OTU0089	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Comamonadaceae	0.98	Acidovorax	0.88	FW301
OTU0637	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Burkholderiaceae	1	Cupriavidus	1	FW301
OTU0314	Actinobacteri	1	Actinobacteria	1	Actinomycetales	1	Micrococcaceae	1	Arthrobacter	1	FW301
OTU0100	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Comamonadaceae	1	Curvibacter	0.88	FW301
OTU0094	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Comamonadaceae	1	Albidiferax	0.79	FW301
OTU0059	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Burkholderiales inc. sed.	0.81	Aquabacterium	0.81	FW301
OTU0176	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Oxalobacteraceae	1	Massilia	0.89	FW301
OTU0196	Proteobacter	1	Gammaaproteobacteria	1	Pseudomonadales	1	Pseudomonadaceae	1	Pseudomonas	0.98	FW301
OTU0661	Proteobacter	1	Betaproteobacteria	0.99	Hydrogenophillales	0.16	Hydrogenophillaceae	0.16	Thiobacillus	0.16	FW301
OTU1028	Proteobacter	0.82	Alphaaproteobacteria	0.6	Magnetococcales	0.31	Magnetococcaceae	0.31	Magnetococcus	0.31	FW301
OTU0045	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	0.99	Burkholderiales inc. sed.	0.81	Aquabacterium	0.79	FW303
OTU0021	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Burkholderiales inc. sed.	0.9	Aquabacterium	0.9	FW303
OTU0279	Proteobacter	1	Betaproteobacteria	1	Rhodocyclales	1	Rhodocyclaceae	1	Azoarcus	1	FW303
OTU0024	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Burkholderiales inc. sed.	0.95	Aquabacterium	0.95	FW303
OTU0964	Proteobacter	1	Alphaaproteobacteria	1	Rhizobiales	1	Methylobacteriaceae	1	Methylobacterium	1	FW303
OTU0719	Proteobacter	0.96	Gammaaproteobacteria	0.95	Oceanospirillales	0.31	Oleiphilaceae	0.18	Oleiphilus	0.18	FW303
OTU0088	Proteobacter	1	Betaproteobacteria	0.99	Burkholderiales	0.99	Comamonadaceae	0.99	Delftia	0.82	FW305
OTU0290	Proteobacter	1	Betaproteobacteria	1	Rhodocyclales	0.97	Rhodocyclaceae	0.97	Dechloromonas	0.57	FW305
OTU0322	Proteobacter	1	Deltaproteobacteria	1	Desulfuromonadales	1	Geobacteraceae	1	Geobacter	1	FW305
OTU0076	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Comamonadaceae	0.99	Acidovorax	0.65	FW305
OTU0477	Elusimicrobia	0.18	Elusimicrobia	0.17	Elusimicrobiales	0.17	Elusimicrobiaceae	0.17	Elusimicrobium	0.17	FW305
OTU0142	Proteobacter	1	Betaproteobacteria	1	Burkholderiales	1	Comamonadaceae	1	Pelomonas	0.62	FW305

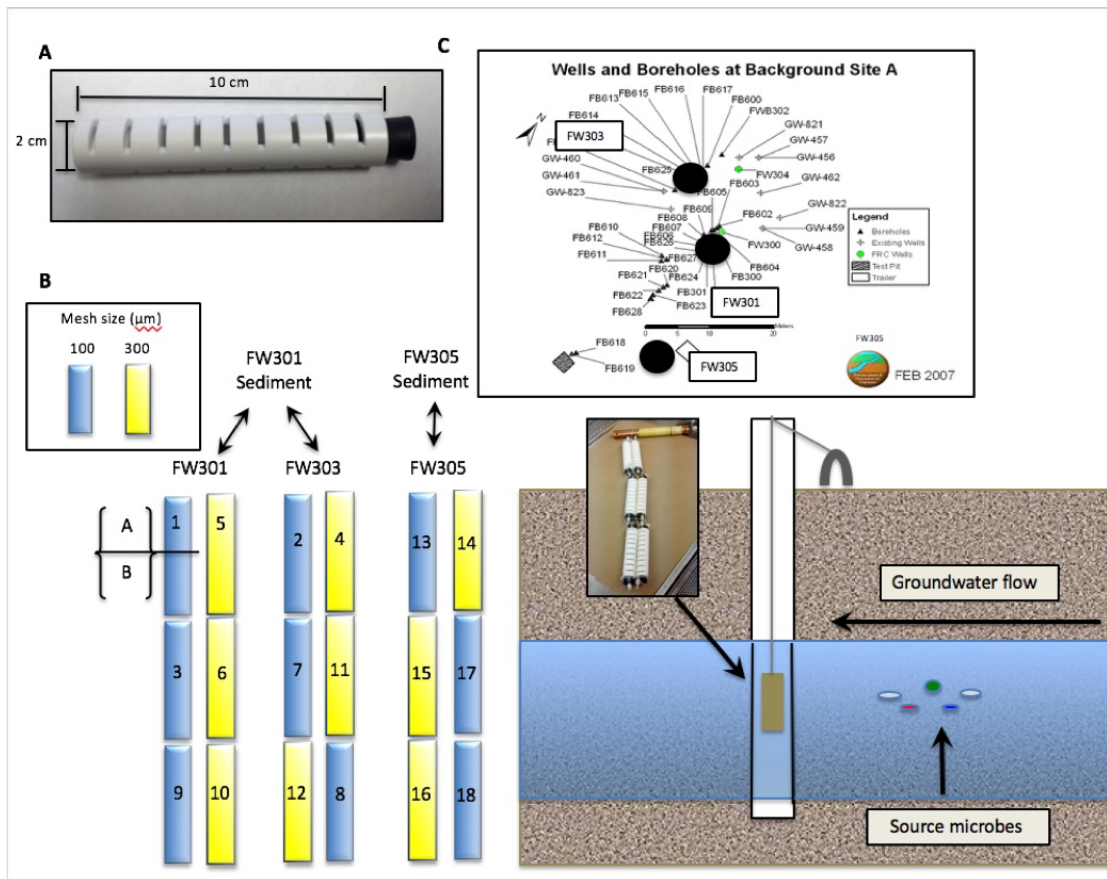


Supplementary Figure A.11. Average pair-wise Bray-Curtis dissimilarities of resident OTUs over time per well. On average, the relative abundances of resident OTUs observed differ such that average Bray-Curtis dissimilarity from day today ranges between 0.49 ± 0.11 and 0.75 ± 0.11 in FW301, 0.44 ± 0.16 and 0.78 ± 0.06 in FW303, and 0.52 ± 0.18 and 0.95 ± 0.05 in FW305.

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER THREE

Supplementary Figure B.1.



A) Bio-trap surrogate sediment samplers comprised of PVC material and capped with a butyl rubber stopper were used for surrogate sediment incubation. B) 18 total bio-traps were filled with either FW301 (1-12) or FW305 (13-18) native sediment prior to autoclaving. Bio-traps were lined with either 100μm (blue) or 300μm (yellow) mesh. Prior to pulverization each bio-trap was cut in half representing duplicates (A and B) for each bio-trap. C) Three field wells were chosen within background site A of the ENIGMA-Oak Ridge Field Research Center.

Supplementary Table B.1. Sequence and OUT counts for Groundwater and Surrogate Sediment samples

	All Data	Groundwater	Surrogate Sediment
Total Sequences (Raw)	1,008,111	601,514	406,597
Total Sequences After Downstream Filtering	990,545	584,652	405,893
Total unique OTUs	3181	3167	1743
Total unique Archaeal OTUs	34	34	12
Total unique Bacterial OTUs	3147	3147	1731

Supplementary Figure B.2 and Supplementary Table B.2.

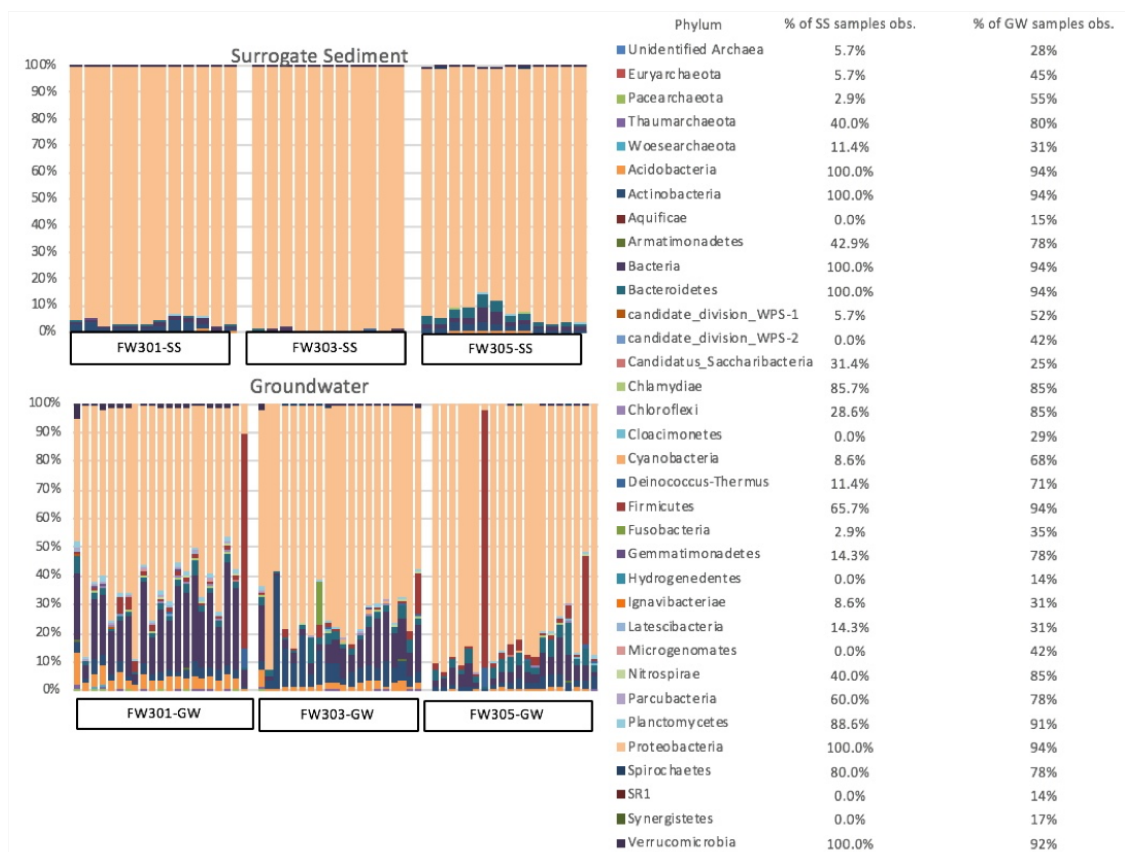


Fig. B.2. Relative Abundance plots of Phyla observed per sample. Table B2. List of sampled Phyla in groundwater and surrogate sediment and proportion of samples in which they were observed

Supplementary Figure B3. Venn-Diagram of shared OTUs A) across groundwater and B) surrogate sediment samples.

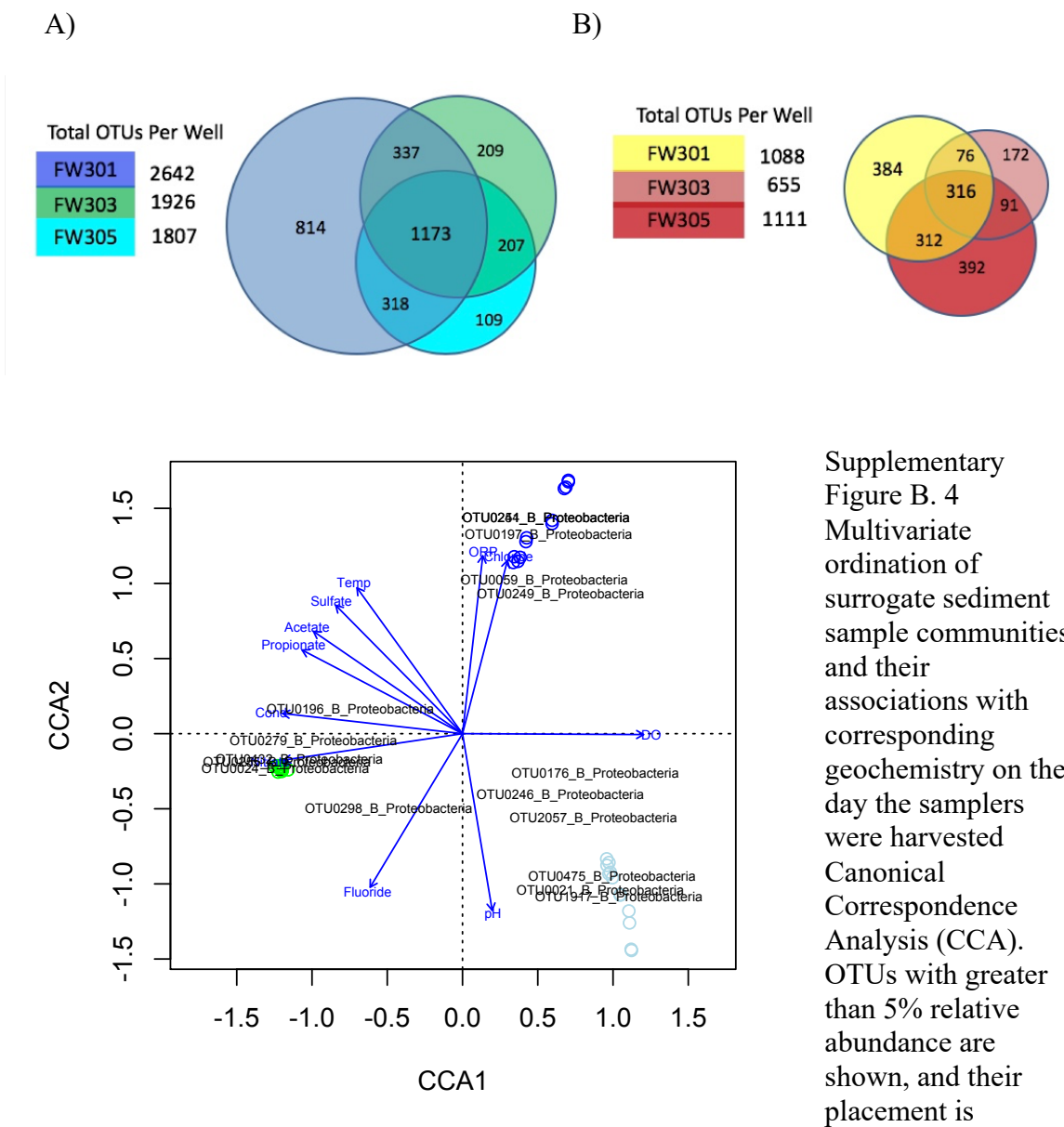


Table B.3. OTUs observed at >5% relative abundance and strongly associated with surrogate sediment samplers on constrained ordination their taxonomic classification based on RDP alignment.

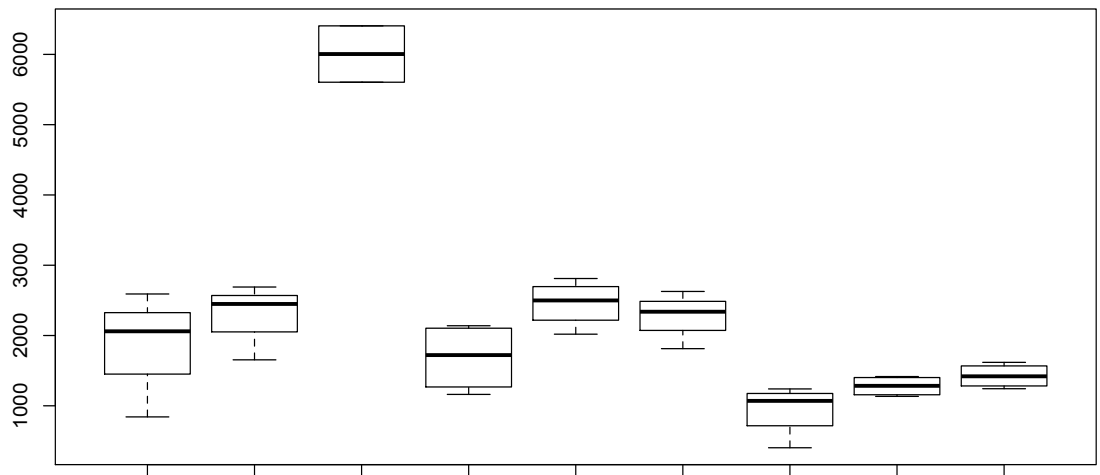
OTUId	Phylum	RDP C.I.	Class	RDP C.I.	Order	RDP C.I.	Family	RDP C.I.	Genus	RDP C.I.	Strongest CCA association
OTU1917	Proteobacteria	1	Alphaproteobacteria	1	Sphingomonadales	1	Sphingomonadaceae	0.96	Sphingobium	0.93	FW301
OTU0244	Proteobacteria	1	Gammaproteobacteria	0.98	Pseudomonadales	0.97	Moraxellaceae	0.97	Perluclidibaca	0.63	FW301
OTU0251	Proteobacteria	1	Gammaproteobacteria	0.98	Pseudomonadales	0.98	Moraxellaceae	0.98	Perluclidibaca	0.68	FW301
OTU0059	Proteobacteria	1	Betaproteobacteria	1	Burkholderiales	1	Burkholderiales_incertae_sedis	0.81	Aquabacterium	0.81	FW301
OTU0249	Proteobacteria	1	Gammaproteobacteria	0.98	Pseudomonadales	0.98	Moraxellaceae	0.98	Paraperluclidibaca	0.59	FW301
OTU0255	Proteobacteria	1	Gammaproteobacteria	0.99	Pseudomonadales	0.99	Moraxellaceae	0.99	Paraperluclidibaca	0.56	FW303
OTU0298	Proteobacteria	1	Betaproteobacteria	1	Burkholderiales	1	Oxalobacteraceae	0.97	Noviherbaspirillum	0.31	FW303
OTU0432	Proteobacteria	1	Betaproteobacteria	1	Rhodocyclales	0.35	Rhodocyclaceae	0.35	Propionivibrio	0.25	FW303
OTU0279	Proteobacteria	1	Betaproteobacteria	1	Rhodocyclales	1	Rhodocyclaceae	1	Azoarcus	1	FW303
OTU0196	Proteobacteria	1	Gammaproteobacteria	1	Pseudomonadales	1	Pseudomonadaceae	1	Pseudomonas	0.98	FW303
OTU0024	Proteobacteria	1	Betaproteobacteria	1	Burkholderiales	1	Burkholderiales_incertae_sedis	0.95	Aquabacterium	0.95	FW303
OTU0176	Proteobacteria	1	Betaproteobacteria	1	Burkholderiales	1	Oxalobacteraceae	1	Massilia	0.89	FW305
OTU0021	Proteobacteria	1	Betaproteobacteria	1	Burkholderiales	1	Burkholderiales_incertae_sedis	0.9	Aquabacterium	0.9	FW305
OTU2057	Proteobacteria	1	Betaproteobacteria	1	Hydrogenophiales	0.31	Hydrogenophilaceae	0.31	Sulfuricella	0.31	FW305
OTU0197	Proteobacteria	1	Gammaproteobacteria	1	Pseudomonadales	1	Pseudomonadaceae	1	Pseudomonas	1	FW305
OTU0475	Proteobacteria	0.98	Gammaproteobacteria	0.94	Alteromonadales	0.17	Pseudoalteromonadaceae	0.14	Algicola	0.14	FW305
OTU0246	Proteobacteria	1	Gammaproteobacteria	0.86	Pseudomonadales	0.64	Moraxellaceae	0.64	Perluclidibaca	0.47	FW305

APPENDIX C

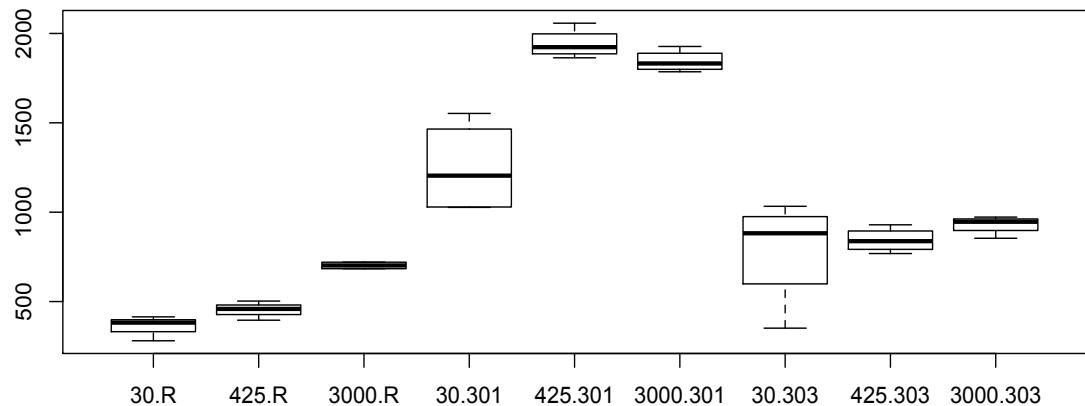
SUPPLEMENTARY MATERIAL FOR CHAPTER FOUR

Supplementary Figure C.1.

A)



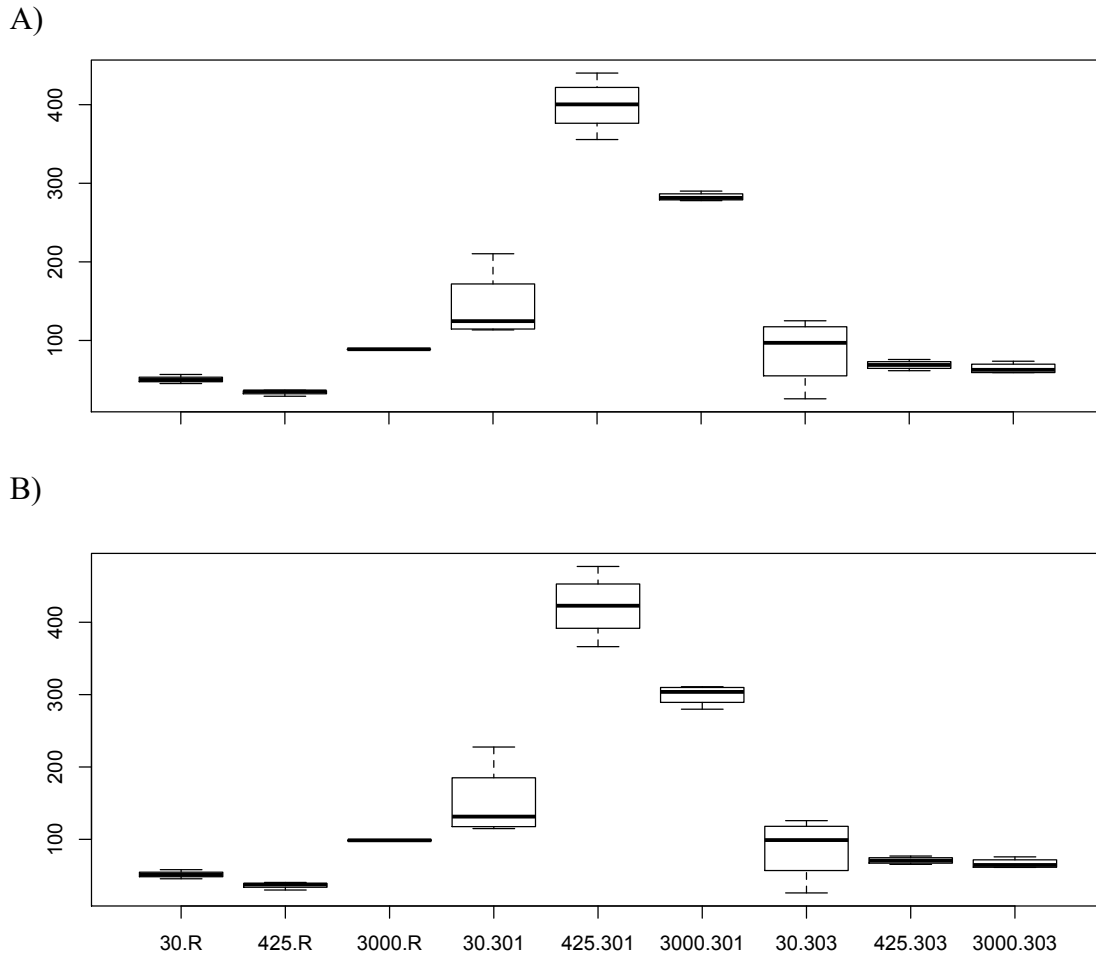
B)



Boxplots of Richness per bead size (30, 425, 3000) per environment (R=Reactor; 301=Field well FW301; 303 = Field well FW303) on A) observed and B) rarefied abundances. A one-way ANOVA was performed and tested with Tukey Honestly Significant Difference test (TukeyHSD). Across all environments of rarefied data, beads incubated within FW301 showed significant differences between beads of all other environment types (in-lab reactors and FW303). Within FW301, differences in richness observed on the smallest sized beads were statistically significantly different than those

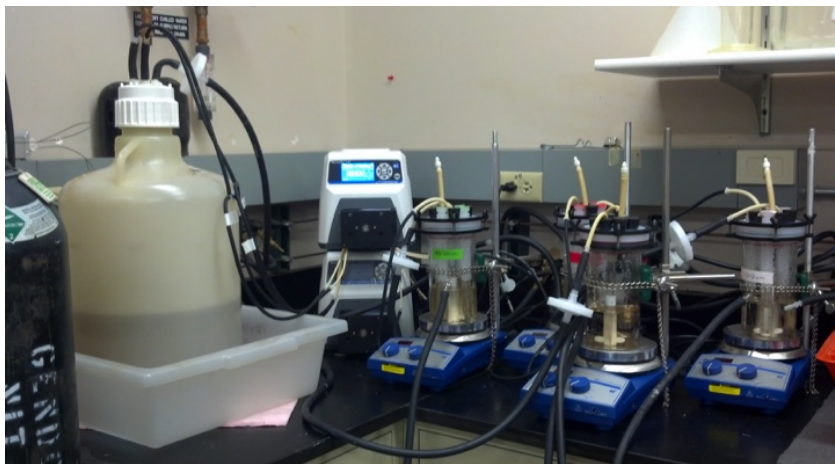
observed on intermediate ($p\text{-value} = 0.00005$) and large ($p\text{-value}=0.0004$) bead. No such differences were observed when comparing beads within in-lab reactors or within field well FW303.

Supplementary Figure C.2.



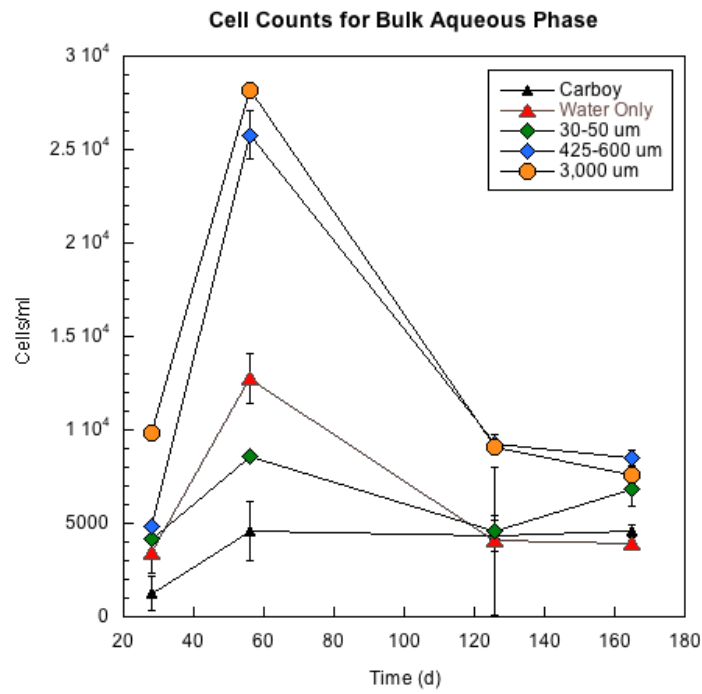
Boxplots of effective species number (Hq1) per bead size (30, 425, 3000) per environment (R=Reactor; 301=Field well FW301; 303 = Field well FW303) on A) observed and B) rarefied abundances. A one-way ANOVA was performed and tested with Tukey Honestly Significant Difference test (TukeyHSD).

Supplementary Figure C.3.

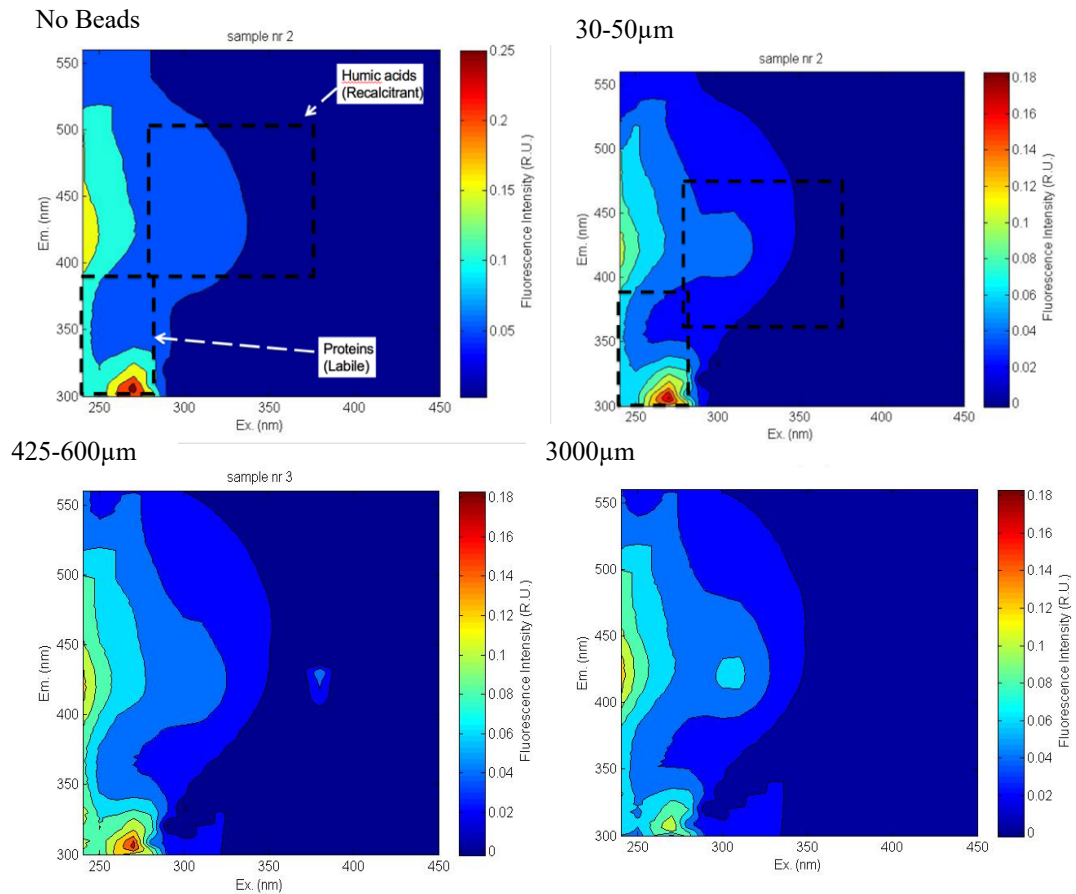
Supplementary Table C.1. Number of beads per draw for 30-50 μm fraction

	Reactor	FW301	FW303
Draw 1	170	85	75
Draw 2	20	109	98
Draw 3	208	124	150
Draw 4	NA	50	35

Supplementary Figure C.4. Total cell counts and organic matter content in laboratory reactors across particle size.



Supplementary Figure C.5



Results of EEAMS analysis based on replicates for effluent collected from each reactor at the end of the six-month sampling period. Graphs indicate areas of high or low intensity based on emission/excitation of organic matter. Samples from no bead control reactors show higher intensity at lower excitation wavelengths (250-300nm), indicating greater concentration of labile organic matter (*e.g.* proteins). Fluorescence intensity shifts can be observed from effluent collected from reactors with increasingly larger bead sizes, indicating compositional shifts in organic matter towards more recalcitrant sources (*e.g.* humic acids).

APPENDIX DPHYSIOLOGICAL CHARACTERIZATION OF ENVIRONMENTAL ISOLATE
DESULFOSPOROSINUS

Many molecular approaches are limited by the lack of cultured isolates. Environmentally relevant cultured isolates for which physiological studies can inform genetic and molecular surveys and allow for a more complete picture of the microbial ecology of microorganisms. It has been argued that an initial objective of environmental studies should be “to obtain a range of isolates from representatives of members of the different phyla and classes and to determine to what extent there is functional and genetic diversity within these groups” in order to “give us an initial overview of the potential roles of different soil bacterial groups and will also enable us to learn the tricks required to culture the more recalcitrant relatives.” (pg. 1726, Janssen 2006).

Towards this aim, sediment samples from EB106 and EB271 were made into a slurry (XXg/xxmL), plated on freshwater media (amended with Mn), and incubated anaerobically at either 23°C or 15°C for three weeks. Once individual colonies appeared they were successively transferred for three generations until isolated. Growth was not evident on any plates incubated at 15°C, rather the four obtained isolates all grew at 23°C. Isolates have been obtained (1)EB106 and (3) EB271.

Of the four isolates we have 16S rRNA sequences for the EB106 isolate and one of the EB271 isolates. BLAST analysis returned *Desulfosporosinus* and *Pelosinus* as the closest genera for EB106 and EB271, respectively. Attempts to grow isolates in liquid

culture (freshwater media and LS4D) at various temperature and agitation conditions (23°C with shaking, 23°C without shaking, 37°C with shaking, 37°C without shaking) were moderately successful, with optical densities (OD600) reaching ~0.2 over approximately a 1 month incubation for all conditions at 23°C. Higher temperatures did not produce growth.

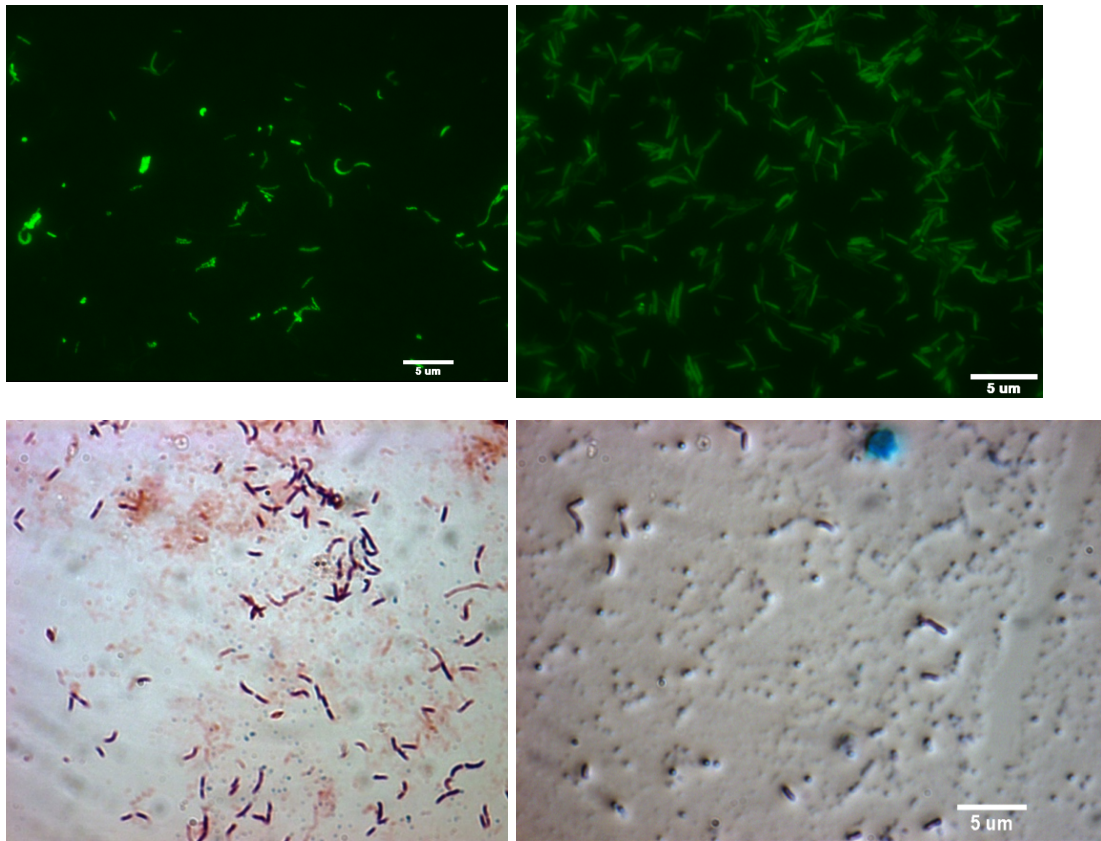


Figure D.1. Microscopic images of D1 isolate. Fluorescent cell staining and cell counts using SYBR Gold (A,B) resulted in an average cell density of 122.6 *cells/mL* taken from OD of X. C) Gram stain reaction showed gram positive to gram variable rods. D) Spore staining using Malachite Green showed positive for spore body formation.

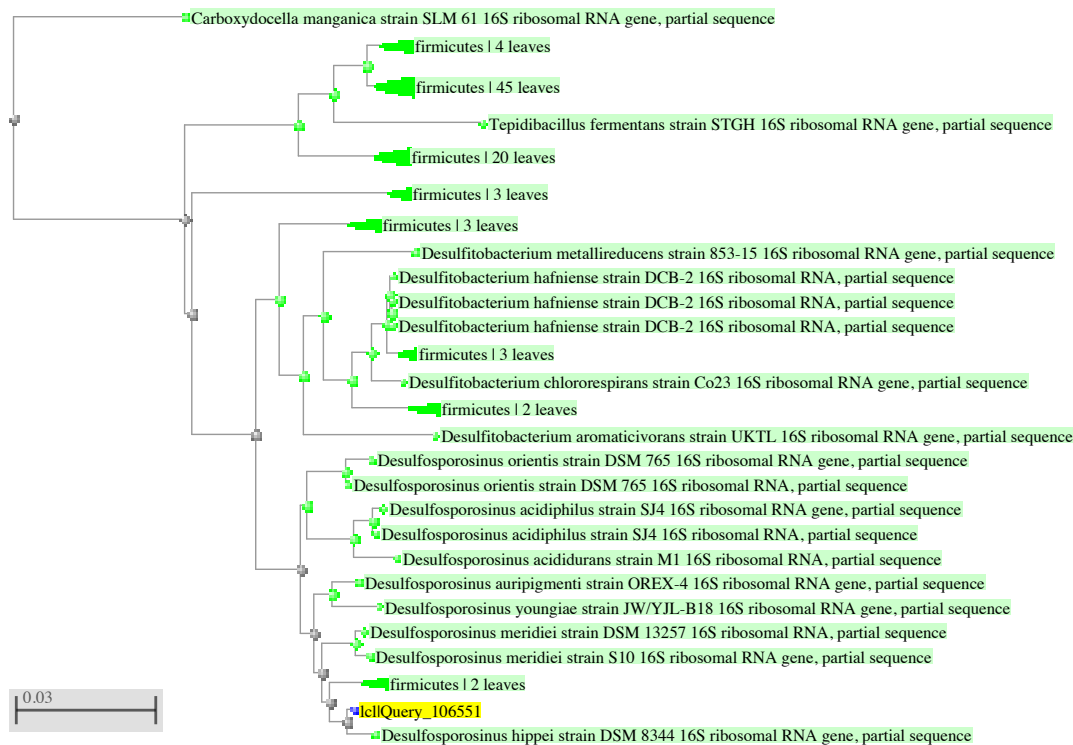


Figure D.2. Phylogenetic relatedness of unknown isolate D1 obtained from sediments cores originating from contaminated core EB106. Based on NCBI-BLAST analysis isolate D1 is most closely related to *Desulfosporosinus* sp.

Figure D.3.

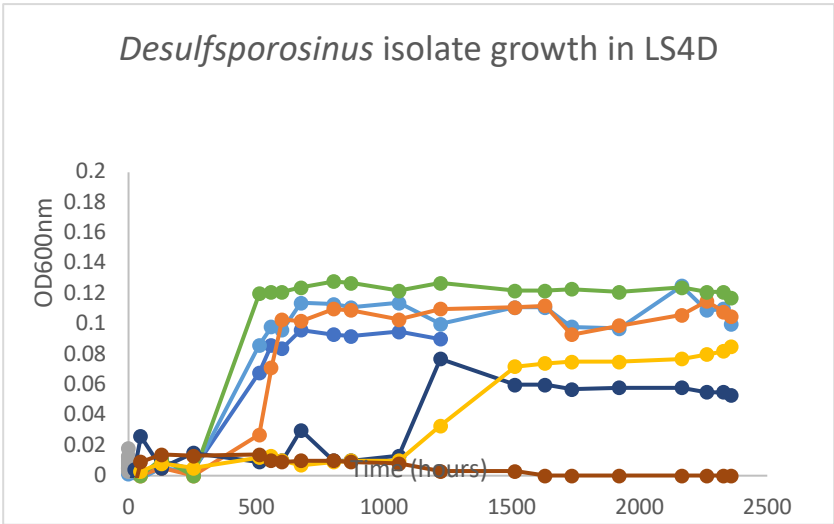


Table D.1. Gram reaction and cell size comparisons for other *Desulfosporosinus* strains relative to the cell's age, and media in which cells were cultivated described in the literature

<u>Name</u>	<u>Gram Reaction</u>	<u>Length of Incubation</u>	<u>Media</u>	<u>Size (µm)</u>
<i>Desulfosporosinus acididurans</i> ⁹	Gram-positive rods that stain gram-negative	-	Anoxic basal media	3-7 in length, and 0.7 in width
<i>Desulfosporosinus acidiphilus</i> ¹³	Gram-variable	3 weeks		4-7 in length, and 0.8 – 1.0 in width
<i>Desulfosporosinus hippei</i> ¹⁰	Gram-negative	-	Basal media amended with sodium lactate and sodium sulfate	4.5-5.5 in length, and 1.0 – 1.2 in width
<i>Desulfosporosinus lacus</i> ¹¹	Gram-negative	-	Basal media amended with 10 mM sodium lactate and 5 mM sodium sulfate	2.3 – 3.8 in length, and 0.55 – 0.7 in width
<i>Desulfosporosinus meridiei</i> ¹²	Gram-negative/Gram-variable	2 weeks	Cultivation medium	2.3-4.2 in length, 0.7-1.1 in width
D1 (2/19/18)	Gram-variable	8 weeks	LS4D media	-
D1 (3/17/18)	Gram-positive	4 weeks	FMnO ₂ media	1.19 – 1.88 in length

Table D.2. Types of media used for growth of D1 isolate

Original media	Modification #1	Modification #2	Modification #3	Modification #4
FMnO ₂ (contains yeast extract, succinate, sulfate, and MnCl ₂)	Media was reduced with either L-cysteine or sodium sulfide	Media was made with different concentrations of ammonium chloride (1M and 2M) and without FeNH ₃ Citrate	Tubes were open to the glove bag's atmosphere using needles and gas filters to allow for gas exchange	-
LS4D (contains 60mM Lactate and 30mM Sulfate)	60 mM Formate and 30mM acetate was added to tubes.	Tubes containing LS4D media were placed in a shaking incubator at 22°C and 150 rpm.	Tubes containing LS4D media were placed in a shaking incubator at 37°C and 150 rpm.	-
CCM (contains	Glucose was added at a concentration of 30 mM	Sucrose was added at a concentration of 30 mM	Both glucose and sulfate were added at a concentration of 30 mM	10 mM glucose, 10 mM fructose, 10 mM pyruvate, 10 mM lactate, and 30 mM sulfate was added.
CSP (contains	-	-	-	-

APPENDIX E

IMPACT OF HYDROLOGICAL BOUNDARIES ON MICROBIAL PLANKTONIC
AND BIOFILM COMMUNITIES IN SHALLOW TERRESTRIAL SUBSURFACE
ENVIRONMENTS

Contribution of Authors and Co-Authors

Manuscript in Appendix D

Author: Heidi J. Smith

Contributions: Performed literature review, synthesized ideas, wrote and edited the manuscript

Co-Author: Anna J. Zelaya

Contributions: Performed literature review, synthesized ideas, wrote and edited the manuscript

Co-Author: Kara Bowen De Leon

Contributions: Wrote and edited the manuscript

Co-Author: Romy Chakraborty

Contributions: Wrote and edited the manuscript

Co-Author: Dwayne Elias

Contributions: Reviewed and edited the manuscript

Co-Author: Terry Hazen

Contributions: Synthesized ideas, reviewed and edited the manuscript

Co-Author: Adam P. Arkin

Contributions: Synthesized ideas, reviewed and edited the manuscript

Co-Author: Albert B. Cunningham

Contributions: Reviewed and edited the manuscript

Co-Author: Matthew W. Fields

Contributions: Principal Investigator, Synthesized ideas, wrote and edited the manuscript

Manuscript Information Page

Heidi J. Smith, Anna J. Zelaya, Kara Bowen De Leon, Romy Chakraborty, Dwayne A. Elias, Terry C. Hazen, Adam P. Arkin, Albert B. Cunningham, Matthew W. Fields

Status of Manuscript:

☐ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☒ Published in a peer-reviewed journal

Published in:

FEMS Microbiology Ecology, 94, 2018, fyy191

doi: 10.1093/femsec/fyy191

ABSTRACT

Subsurface environments contain a large proportion of planetary microbial biomass and harbor diverse communities responsible for mediating biogeochemical cycles important to groundwater used by human society for consumption, irrigation, agriculture and industry. Within the saturated zone, capillary fringe and vadose zones, microorganisms can reside in two distinct phases (planktonic or biofilm), and significant differences in community composition, structure and activity between free-living and attached communities are commonly accepted. However, largely due to sampling constraints and the challenges of working with solid substrata, the contribution of each phase to subsurface processes is largely unresolved. Here, we synthesize current information on the diversity and activity of shallow freshwater subsurface habitats, discuss the challenges associated with sampling planktonic and biofilm communities across spatial, temporal and ecological gradients, and discuss how biofilms may be constrained within shallow terrestrial subsurface aquifers. We suggest that merging traditional activity measurements and sequencing/-omics technologies with hydrological parameters important to sediment biofilm assembly and stability will help delineate key system parameters. Ultimately, integration will enhance our understanding of shallow subsurface ecophysiology in terms of bulk-flow through porous media and distinguish the respective activities of sessile microbial communities from more transient planktonic communities to ecosystem service and maintenance.

Introduction

The terrestrial, shallow subsurface is a complex and microbially-active habitat located beneath the surface soil layers, comprised of sediments (inorganic or organic unconsolidated material that comes from the weathering of rock transported by wind, water or ice), rocks, gas, pore water and groundwater (Atekwana, Werkema and Atekwana 2006). Typically, subsurface environments contain less labile organic matter (OM) compared to surface soils, and the degree of hydrological connectivity to the surface is routinely used to delineate between shallow and deep biospheres rather than depth alone (Lovley and Chapelle 1995). Although water covers 70% of the Earth's surface, roughly 1% is readily available for human use, and a vast majority (~95%) of the Earth's consumable and available freshwater is groundwater (Danielopol *et al.* 2008; Griebler *et al.* 2014; Dennehy, Reilly and Cunningham 2015). Despite the importance for the world's population, the role of microbial communities in the maintenance of groundwater ecosystems is not fully understood. Case in point, the recent increase of artificially recharging natural aquifers via managed aquifer recharge to meet the global demand for water availability is concerning because of the potential to drastically alter groundwater systems (Lee and Lee 2017). This mini-review will focus on aspects of 'shallow' freshwater subsurface environments (mainly porous/granular) which typically have higher rates of recharge and flow as well as have a high degree of connectedness with the surface as opposed to 'deep' subsurface environments that are much less connected with the surface and receive limited surficial inputs of water and/or nutrients.

Primary motivations for studying the subsurface are to expand what is known about Earth's microbial diversity and the subsurface microorganisms under low nutrient conditions that significantly impact C, S, N, P and mineral cycles. Microbial life is thought to vary from the terrestrial surface to the deep subsurface dependent upon water, nutrient inputs and environmental stressors. Upwards of 40% of the microbial biomass and 10^{16} – 10^{17} g C on Earth resides within the terrestrial subsurface (Whitman, Coleman and Weibe 1998; Griebler and Lueders 2009; McMahon and Parnell 2014). Over the last 30 years, there has been an increasing interest in surveying the taxonomic and functional biodiversity of subsurface environments, largely due to the concern over biodiversity and subsequent ecosystem function loss (Danielopol *et al.* 2003; Hancock, Boulton and Humphreys 2005; Wall and Nielsen 2012; Lijzen, Otte and van Dreumel 2014). However, on an ecosystem scale, there is limited information regarding the exact relationship between microbial diversity, environmental parameters and biogeochemical processes between groundwater and subsurface porous media. Studies focusing on subsurface habitats have revealed many significant roles that microorganisms play in shallow subsurface processes (*e.g.* Chapelle 2000; Atekwana, Werkema and Atekwana 2006; Hwang *et al.* 2009; Mitchell *et al.* 2010; Akob and Kusel 2011; Griebler and Avramov 2015).

In the environment, microorganisms can be observed in two distinct phases: free-living (planktonic) and associated with a surface as single cells to multicellular aggregates (*i.e.* biofilm). Biofilms are often composed of diverse taxonomic lineages attached to surfaces and each other, typically surrounded by extracellular polymeric substances (Hall-Stoodley, Costerton and Stoodley 2004; Gross *et al.* 2007; Stewart and Franklin 2008).

Biofilms have not been explicitly studied within the subsurface; however, because biofilms have been described at liquid–solid, liquid–gas or solid–solid interfaces, it is becoming increasingly clear that biofilms more closely resemble *in situ* conditions for microorganisms from diverse environments (Hall-Stoodley, Costerton and Stoodley 2004). Therefore, it is likely that attached modes of growth are a universal feature presenting an important physiology to explore within the subsurface in addition to typically conducted planktonic cell studies (Dunne 2002; Kolter 2005).

Cells growing on a surface (*i.e.* biofilms) are known to have physiologies and properties distinct from planktonic cells including increased resistance to external stresses such as antimicrobials, heavy metals, desiccation and substrate deprivation (*e.g.* Clark *et al.* 2012; Kurczy *et al.* 2015; Stylo *et al.* 2015). Most microbial environments are physically dynamic habitats where fluxes in water, nutrients, temperature, pH and osmolarity can create challenges for survival. Altered flow conditions can limit motility/dispersal and nutrient availability can result in decreased microbial activity and altered population distribution (Or *et al.* 2007). Biofilm matrices can retain water, sorb nutrients and protect against rapid changes in local geochemistry, attributes that significantly improve microbial viability and activity. Additional ecophysiological advantages from residing within biofilms include metabolic cooperation, the exchange of genetic material and the development of regulatory mechanisms and social behaviors (Dang and Lovell 2016 and references therein).

Traditionally, subsurface habitats were analyzed through bulk activity assays and total and viable cell enumerations (mainly with groundwater samples). Recent studies have relied on sequencing and -omics techniques to identify new diversity and functionality. Unfortunately, little overlap exists between more traditional quantitative activity

measurements and newer sequencing capabilities. Such overlap is necessary to link phylogenies to quantitative functionality, although systems approaches have been used for bioremediation sites (Chakraborty, Wu and Hazen 2012). The objectives of this review are to synthesize the current understanding of (i) microbial population distributions and activities spanning shallow subsurface habitats (with a focus on freshwater systems when possible), (ii) discuss the challenges associated with sampling planktonic and biofilm communities across spatial, temporal and geological gradients, (iii) identify subsurface geochemical and physical properties that potentially constrain biofilm development and iv) give recommendations and considerations for future studies. Additionally, Table S1 (Supporting Information) provides a quick summary of sample and environmental details, including lithologic information when available, for the relevant sources cited in this review.

Characteristics of Shallow Subsurface Environments

Although estimates vary, the shallow subsurface environment can extend from beneath the OM rich soil layers (A and B horizons) to tens of meters (Atekwana, Werkema and Atekwana 2006; Pepper and Brusseau 2006). In the shallow subsurface environment of an aquifer, sediments are assumed to lie below the vertically weathered top soil profiles. In the B horizon (below O, A and E), minerals, clays and organic material are leached from the upper horizons. The C horizon (below B) is characterized by unweathered minerals that were the parent material from which the upper soils were formed while at deeper depths the R horizon is the native bedrock material (Pepper and Brusseau 2006). The shallow subsurface is typically described as being below the surface soil horizons (typically 1–10 cm) and above bedrock (<50 m in depth) (Chu *et al.* 2016), and can have a high degree of hydrological connectedness with the surface compared to the deep subsurface (Toth 1963; Lovley and Chapelle 1995). By contrast, deep subsurface systems have been distinguished by arbitrary depth measurements ranging from hundreds to thousands of meters below the surface (Balkwill 1989; Lovley and Chapelle 1995; Head, Jones and Larter 2003) or by a lack of surface connectivity (Toth 1963; Lovley and Chapelle 1995). The designation of

‘shallow’ versus ‘deep’ can be variable dependent upon respective geology and environment. Additionally, albeit not within the scope of this minireview, there are different concepts to categorize aquifers (*e.g.* aquifer, aquiclude and aquitard) in the context of hydrology.

Traditionally, the shallow subsurface can be separated into three distinct zones based on moisture content in relationship to water table configuration termed the vadose, capillary fringe and saturated zones (Fig. 1). The vadose zone represents the upper most boundary of the subsurface comprised of the upper horizons (O–B) and contains unweathered and weathered materials. Following precipitation events, the vadose zone experiences high saturation levels as vertical infiltration proceeds downward to the water table, yet residual pore water can persist creating varying levels of water and gas saturation (Jones and Bennett 2014). The capillary fringe exists at the interface of the saturated and vadose zone and is highly dependent upon fluctuations of the local water table. The capillary fringe is dynamic over time with varying physicochemical conditions resulting from water table fluctuations (Griebler and Lueders 2009). This fluctuating interface has been shown to be a ‘hotspot’ of subsurface activity especially with respect to biogeochemical cycling (Silliman *et al.* 2002; Berkowitz, Silliman and Dunn 2004). The saturated zone (*i.e.* at/below water table) of most aquifers consists of porous parent material (C and R horizons) and voids are filled with water. Generally, the direction of water flow in the saturated zone can be 3-dimensional depending on hydraulic gradients and porous media properties (*e.g.* clay lenses). With respect to the impact on microbial communities,

much attention has been given to the water table position and sediments transported in the saturated and capillary fringe zones.

The latter transitional boundary between the vadose and saturated zones is capable of drastic changes in geochemical parameters [*e.g.* pH and dissolved oxygen] that impact ecosystem function in terms of geochemical cycling, biotic/abiotic filtering and buffering processes (Rainwater *et al.* 1993; Reddi, Han and Banks 1998; Dobson, Schroth and Zeyer 2007; Pilloni 2011; Chakraborty, Wu and Hazen 2012). However, the vadose zone can also have soils and weathered particles (sediments) impacted by water movement, likely dictated by surface infiltration and evapotranspiration. In addition, clays and clay lenses are also thought to impact water and gas flow that could significantly impact microbial processes (Faybishenko *et al.* 2000). In addition, particle structure can impact community composition and activity, for example, the turnover of matrix-associated NOM (natural organic matter) correlates to the proportion of fine-grained particles (Keil and Mayer 2014). Physical properties of sediments (*e.g.* particle size) is also thought to impact microbial activity and distribution although most studies have been done with surface or

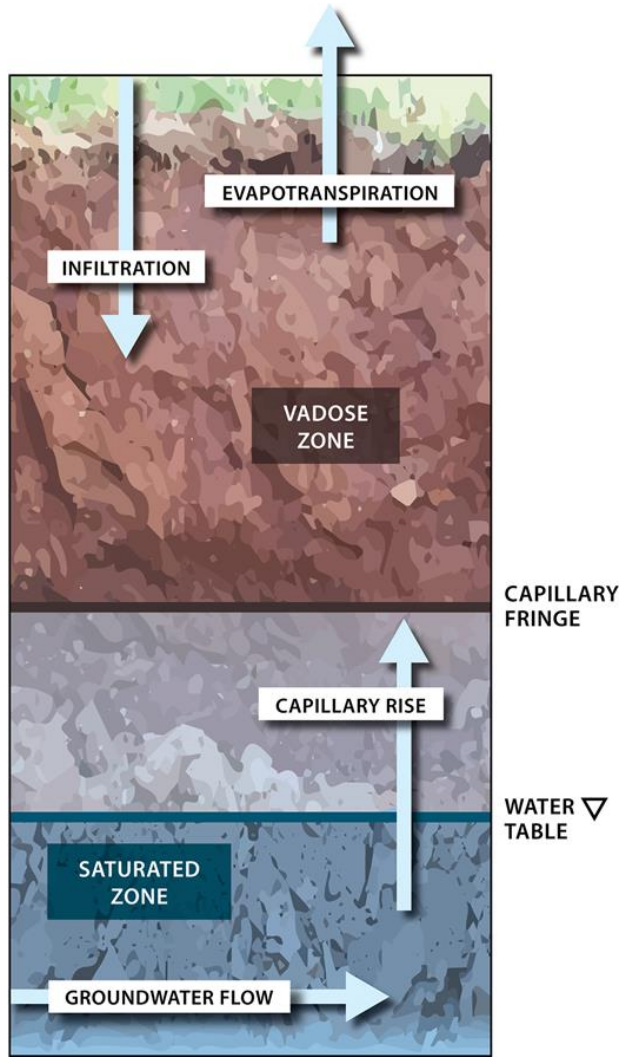


Figure E.1. Conceptual illustration of representative shallow subsurface environment that includes the vadose, capillary fringe, and saturated zones. Arrows depict the movement of water through infiltration, evapotranspiration, capillary rise and re-charge, and the movement of water within and between these zones creates dynamic conditions for the formation and maintenance of subsurface biofilms.

near-surface soils/sediments (*e.g.* Jackson and Weeks 2008; Hemkemeyer *et al.* 2015, 2018). Individual aggregates in groundwater and soil as well as soil pores can have discrete microenvironments with distinct activities and conditions (Keil and Mayer 2014) that likely contribute to spatial and temporal areas of high metabolic activities or ‘hot spots and hot moments’ (McClain *et al.* 2003). Pores and aggregates are continuously changing due to biogeochemical and physical processes (Schlüter and Vogel 2016), and wetting/drying

cycles (*i.e.* capillary fringe) can greatly impact pore size distributions (Bodner, Scholl and Kaul 2013). Sediment–groundwater– cell interactions can occur at the pore scale ($\leq$ micrometer) where diffusion and dispersal can be limited. However, little is known about how microbially relevant scales ultimately impact field scale behavior and function, and few studies have determined the proper scale to delineate these relationships.

While the subsurface begins below the humus rich soil horizons, NOM (including particulate and dissolved fractions excluding organic contaminants) is a primary source of C/N that supports microbial life in the shallow subsurface. Despite seasonal shifts, there is a natural gradient of decreasing nutrient and oxygen concentrations with depth leading to oligotrophic and anoxic conditions within the saturated zone (Danielopol, Pospisil and Rouch 2000; Awoyemi, Achudume and Okoya 2014). Additionally, NOM is thought to decline with depth, and recent comparisons of water-extractable organic matter from a shallow subsurface core showed total organic carbon was ~ 19 mg/g and inorganic carbon was 8 mg/g in shallow sediment (Chakraborty *et al.*, unpublished data). Due to nutrient limiting conditions, microorganisms in subsurface habitats have most likely developed strategies to use NOM and other reduced compounds (*e.g.* Mn(II), Fe(II), ammonia, sulfide, methane and hydrogen) as part of directly or indirectly coupled processes in the groundwater, pore water and sediment surfaces.

Shallow Subsurface Microbial Biodiversity: Progress and Challenges

The relationship between biodiversity and ecosystem functioning has been well studied above ground (Cardinale *et al.* 2006; Ives and Carpenter 2007); however, similar studies are in the early stages for subsurface environments. While perceived functional redundancy

could have a limited role in subsurface ecosystem functioning, studies also indicate that microbial taxonomic diversity plays a role in mitigating ecosystem collapse and contributing to faster functional recovery (Wagg *et al.* 2014; Delgado-Baquerizo *et al.* 2016; Louca and Doebeli 2016). Subsurface groundwaters and sediments have been shown to harbor far more taxonomic and functional diversity than previously inferred by cultivation attempts and microscopic observations (Brown *et al.* 2015; Lynch and Neufeld 2015; Lennon and Locey 2016). In addition, these environments exhibit a wide diversity of previously undescribed bacteria and archaea (Castelle *et al.* 2013; Brown *et al.* 2015; Anantharaman *et al.* 2016; Lazar *et al.* 2017). While specific taxonomic lineages can be prevalent in several types of underground ecosystems (Griebler and Lueders 2009; Akob and Ku"sel 2011; Hubalek *et al.* 2016), thus far no true 'endemic' shallow subsurface populations have been identified (Griebler and Lueders 2009). To date, the debate regarding the influence of biodiversity and ecosystem functioning, especially within the subsurface, has yet to be thoroughly explored. Any resolution will most likely be challenging at best due to extreme spatial heterogeneity.

It remains unresolved whether our current understanding of subsurface microbial biodiversity is real or merely an artifact of the following topics: (i) technological approaches (*i.e.* short reads lengths from next generation sequencing), (ii) low relative diversity and/or abundances of oligotrophic systems, (iii) the use of bulk sampling techniques compared to the retrieval of samples representing discrete phases (planktonic vs. biofilm) and/or discrete zones (*i.e.* vadose, capillary fringe, and saturated zones) which could further delineate spatial differences, and/or (iv) temporal dynamics that have been

poorly resolved. For example, recent work has shown the potential importance of microbial biomass for protozoan food webs in shallow aquifers (Hutchins *et al.* 2016) and differences in carbon cycling between groundwater and shallow sediments over time and space (<1 m) (Longnecker and Kujawinski 2013). Given these types of observations, the roles of biofilm diversity in the shallow subsurface for resistance to predation pressures and ultimately on resource allocation are not known. Therefore, as discussed below, future studies should combine technological approaches at appropriate temporal and spatial scales for both groundwater and matrix material

Technological approaches

Studies of microbial biodiversity have historically been performed via traditional microbiological techniques (Goldscheider, Hunkeler and Rossi 2006 and references therein; Sinclair and Ghiorse 1989). Profound advancements have been made in the application of next-generation sequencing (Tringe and Hugenholtz 2008), high-throughput -omics approaches (López-García and Moreira 2008; Prosser 2015), single-cell methods (Lasken and McLean 2014) and methods encompassing untargeted functional potential (López-García and Moreira 2008; Rajendhran and Gunasekaran 2011). Up until July 2015, a total of $\sim 1.4 \times 10^6$ and $\sim 5.4 \times 10^5$ full-length bacterial and archaeal 16S rRNA reference sequences, respectively, have been deposited into Silva-ARB (www.arb-silva.de) and IMG (img.jgi.doe.gov), comprising a total of 65 bacterial and 20 archaeal phyla (Schloss *et al.* 2016). Interestingly, it was estimated that only 7.8% and 16.5% of all reference sequences originated from soil and aquatic environments, respectively (Schloss *et al.* 2016). As the above estimates include surface waters (*e.g.* lakes and rivers), marine environments and

surface soils, the percentage of sequences specific to groundwater, and more so for shallow subsurface sediments, is quite low. The drastic under-sampling of the subsurface has led to a scarcity of reference sequences specific to these environments, leading to the high risk of mis-identification of retrieved sequences and an under estimation of subsurface biodiversity and biochemical capacity. As the number of non-targeted (DNA/RNA-based) and targeted metagenomes (SIP/activity) are increased for shallow subsurface groundwater and sediments, it is likely that unique lineages with novel capability will be discovered across all three domains, and thus an improved representation of *in situ* diversity can be achieved.

Low relative diversity/abundances of oligotrophic systems

Based upon a limited number of studies that survey diversity as a function of depth, it has been observed that species richness declines over depth, with transient increases at transition zones. Currently, it remains unclear if this trend is merely a consequence of the combination of limited/recalcitrant resources (C) and energy restriction (anoxic) or other more specific selection mechanisms that may differ from surface environments (Musslewhite *et al.* 2003; Lin *et al.* 2012b; Chu *et al.* 2016). Recent studies suggest that the large fraction of lowly abundant or ‘rare’ organisms observed in subsurface environments may play important ecological roles. For example, they may contribute to biogeochemical reactions (Pester *et al.* 2010) while also serving as a ‘microbial bank’ that can ‘seed’ environments when conditions change (Lynch and Neufeld 2015). Biofilms could play a major role across the shallow subsurface zones in which changing conditions (*e.g.* pH, conductivity and flow) could drive dispersal and/or invasion (discussed below).

As oligotrophy is inherent to most subsurface systems, techniques that couple high-throughput manipulation with small volumes and -omics methodology (*e.g.* micro-droplet fluidics and flow cytometry) should be included in future work to enable cultivation and activity measurements of slower growing microorganisms (*e.g.* Wilkins *et al.* 2014).

Spatial variability: Diversity of discrete zones

While mechanisms that affect population distributions have been formulated based on surface habitats (*i.e.* biotic interactions, dispersal limitation and environmental filtering) (Martiny *et al.* 2006; O'Malley 2007; Griebler and Lueders 2009; Shoemaker, Locey and Lennon 2017), it remains unclear whether these mechanisms hold true for the distribution of microorganisms within the oligotrophic subsurface (Musslewhite *et al.* 2003; Chu *et al.* 2016). It is hypothesized that transition zones (macro- as well as micro-transition zones, such as between individual particles and surrounding pore water) are important ecotones or 'hotspots' of microbial diversity and activity (Zhang *et al.* 1998; McClain *et al.* 2003; Goldscheider, Hunkeler and Rossi 2006; Bougon *et al.* 2012; Campbell *et al.* 2012; Jones and Bennett 2014) and deserve more careful attention. There is evidence of spatial (vertical and horizontal) taxonomic variation of groundwater (Lin *et al.* 2011; Lin *et al.* 2012a; Herrmann *et al.* 2015) and sediments (Lin *et al.* 2012b). Results typically show a decline in microbial richness and diversity over vertical depth. The extent that microbial communities vary in relation to depth, even with application of newer sequencing technologies, is still poorly resolved for the variety of geological strata that represent the shallow subsurface. Therefore, increased spatial resolution is needed to better understand the implications of micro-scale heterogeneity on microbial population distributions.

It is not known if observed variation is a consequence of geophysical, geochemical or hydrogeochemical constraints, or a combination thereof. Whereas various scales have been surveyed (cm, m, km) when measuring spatial β -diversity of microbial communities, studies that span over several cm are more common, with deeper samplings that span meters being less frequent. Pronounced effects of horizontal spatial dissimilarity on β -diversity increasing with depth have been shown for surface and subsurface soils (Chu *et al.* 2016), suggesting that, at least down to the saturated zone, subsurface sediment biofilms could be more greatly affected by dispersal limitation than communities of surface soils. Moreover, the proper scale at varied spatial resolution to capture microscale heterogeneity or the proper scale for different geologic strata is currently unknown.

Spatial variability: Diversity of discrete phases

In order to further investigate microbial diversity in the subsurface, it is essential to differentiate between planktonic and attached populations. Historically, the ease of groundwater sampling via well-pumping has resulted in the majority of subsurface datasets. However, inferences made about subsurface communities based solely on the planktonic fraction may not adequately represent all microbial members of the subsurface ecosystem (Hug *et al.* 2015).

Studies that have attempted to compare planktonic versus biofilm communities have resorted to the use of surrogate sediments (native and/or artificial material incubated *in situ* down well) that represent the geology of the aquifer (Reardon *et al.* 2004; Flynn, Sanford and Bethke 2008; Flynn *et al.* 2012; Converse *et al.* 2015; Graham *et al.* 2017).

These surrogates include laboratory microcosms (Lee and Lee 2017), in field biofilm reactors (King *et al.* 2017; Christensen *et al.* 2018), or sediment fines from backwashed pumps (Cardenas *et al.* 2008; Li *et al.* 2018). Early studies that compared the planktonic versus attached fractions have generally observed a subset of the planktonic community in the attached fraction (Hazen *et al.* 1991). Several studies have corroborated these findings over the years (Reardon *et al.* 2004; Brad *et al.* 2008; Anneser *et al.* 2010; Zhou, Kellermann and Griebler 2012).

Studies to differentiate planktonic versus biofilm functions may be able to capture the transitional states of planktonic communities (from planktonic to biofilm and vice versa), but there are unique limitations to each approach. For example, samplers (*e.g.* sampling coupons) typically contain unconsolidated sediments that may not accurately mimic the hydrological effects of consolidated or saturated sediments. Thus, borehole artifacts must be considered (Lehman 2007a). In addition, colonization of the native matrix material is dependent upon surrounding groundwater/porewater. The colonization no doubt occurs *in situ*, but the studies are over short time periods compared to *in situ* conditions and may not achieve the diversity of the natural setting. However, the down-well incubations of solid material does enable the capture of some microbial populations typically missed by groundwater sampling and could capture interaction dynamics across the aqueous/solid matrix boundary under *in situ* conditions (Barnhart *et al.* 2013).

We recently used a revised microbial sampler (patent pending) in a coal-bed aquifer packed with native coal material incubated down-well for ~3 months and compared the bacterial communities (SSU rRNA gene libraries) between sampled groundwater, native

coal core and coal material from the same formation incubated in microbial samplers (Schweitzer *et al.*, unpublished data). Preliminary analyses suggest that some family-level operational taxonomic units were common to all three samples while other operational taxonomic units were common to groundwater and the surrogate matrix material ($n = 3$). Two operational taxonomic units were unique to the sampled groundwater and four unique to the surrogate matrix material. Not surprisingly, the native coal material had the most unique operational taxonomic units ($n = 6$). Whereas differences between the samples were expected, surrogate matrix material could be used in future studies to capture additional diversity from the subsurface and delineate ecology dynamics in terms of core (*i.e.* consistent) and transient populations between groundwater and matrix material.

The inability to distinguish between different phases (*i.e.* groundwater vs. sediments) for key biogeochemical processes poses a challenge for answering basic ecological questions. Recent -omics approaches applied to samples from the subsurface are revealing the range of possible activities within the subsurface and the potential for broad biochemical functionalities (Griebler and Lueders 2009; Akob and Kuşel 2011; Flynn *et al.* 2013; Hubalek *et al.* 2016). Through genomic surveys, microorganisms have been linked to the transformations of carbon compounds, the nitrogen cycle and sequestration of greenhouse gases (Hemme *et al.* 2015; Trivedi, Delgado-Baquerizo and Trivedi 2016). Typically, functionality has been inferred from the presence of specific functional genes in a sample set (*e.g.* Yan *et al.* 2003; Fields *et al.* 2006; Winderl, Schaefer and Lueders 2007) or, more holistically, across a subsurface ecosystem. Usually the larger scale sampling needed to characterize the potential microbial function in an ecosystem relies on testing the

more easily accessible groundwater (Hug *et al.* 2015; Smith *et al.* 2015; He *et al.* 2018). Therefore, despite any novel and potential activity identified in groundwater, the full extent of possible differences in the sediments are unknown. Due to the challenges of getting intact sediment samples, previous studies have used surge blocks (Wu *et al.* 2013) on groundwater pumps to collect sediment fines. This procedure allowed delineation of the groundwater populations from those associated with the sediment fines (Cardenas *et al.* 2008; Li *et al.* 2018), and the reported results showed definite differences between the groundwater and sampled sediments. The potential differences in functionality between the planktonic and sediment microorganisms have implications (*e.g.* selection pressures and dispersal) for ecosystem stability and resiliency, particularly with dynamic hydrological cycles that can change over varying time-scales.

Effects of temporal fluctuations on biodiversity

Taxonomic diversity in relation to spatiotemporal fluctuations in physicochemical and geochemical properties of the shallow subsurface is largely understudied. Physicochemical and geochemical relationships over time may drive phylogenetic and functional microbial diversity changes that are often associated with both short-term (*e.g.* diel cycles; extreme weather) and long-term (*e.g.* seasonal) alterations. Spatiotemporal fluctuations in conjunction with porosity and permeability modifications can lead to varying degrees of hydrogeochemical mixing. The role of mixing in aquatic systems has been implicated in the creation of patchy distributions of both nutrients and biomass (Ebrahimi and Or 2016; 2018). Depending on the aquifer system, highly mixed waters can

be found in shallow areas, where there are higher rates of infiltration from rain or surface waters. Alternatively, at the water table, seasonal fluctuations can result in mixing groundwater with sediments of the variably saturated zone (Fig. 2). In the shallow subsurface, mixing is thought to cause instability due to faster and shorter local flow paths; whereas, more stable and predictable diversity may result from slower regional flow significantly below the water table (Ben Maamar *et al.* 2015).

Most spatiotemporal studies of subsurface environments focus on the changes of geochemistry and corresponding groundwater communities collected via sampling wells over time across different depths in the water table (Lin *et al.* 2012a; Brad *et al.* 2013; Sirisena *et al.* 2013; Ben Maamar *et al.* 2015; Schwab *et al.* 2016). Studies have consistently shown that mixed groundwaters have higher diversity and variability than groundwater that undergoes less mixing (Bougon *et al.* 2012; Hug *et al.* 2015; Danczak *et al.* 2016; Hubalek *et al.* 2016). Additionally, it has been shown that the degree of hydrogeochemical mixing can greatly impact microbial assemblage compositions due to the influx of nutrients and migration of transient populations (Haack *et al.* 2004; Fields *et al.* 2006; Hwang *et al.* 2009; Velasco-Ayuso *et al.* 2009; Lin *et al.* 2012a; Hubalek *et al.* 2016).

The impact of mixing on *in situ* sediment biofilms is largely unknown, mainly because temporal sampling in the same location is nearly impossible for native matrix material. Surrogate sediments incubated down-well have enabled some degree of temporal sampling of attached microorganisms (Zhou, Kellermann and Griebler 2012), as well as successional events when monitored longer times. Comparisons of cataloged cored

sediments separated by months or years have also been performed (Hug *et al.* 2015). It is likely that feedbacks between resident microorganisms and hydrogeochemistry exist and impact subsurface ecosystem structure and function at a larger-scale (Mendoza-Lera and Mutz 2013; Lee and Lee 2017) particularly for sediment biofilms where impacts at the microscale are more likely compared to bulk-phase changes.

Activity in the Shallow Subsurface

There are significantly fewer studies that have simultaneously compared microbial activities in the sediment and groundwater fractions. Quantifying activity from subsurface samples is a non-trivial task, as the retrieval of ‘undisturbed’ samples in combination with ‘representative’ incubation times necessary for activity assays can potentially lead to artifacts which greatly influences downstream analyses and interpretations. While sequencing capabilities have produced substantial insight about the potential functionality of porous subsurface aquifers, traditional -omics studies struggle to make quantitative estimations about activity (Hemme *et al.* 2015; Hug *et al.* 2015). Subsurface activity is typically measured utilizing traditional approaches (*e.g.* extra-cellular enzyme assays, radioisotope tracers, viable plate counts, most probable numbers and direct counts with fluorescent compounds indicative of activity), all of which have been shown to have inherent biases (Kepner and Pratt 1994; Stewart *et al.* 1994). Historically, not only were there greater densities of total cells in the sediments, but a higher proportion of active cells

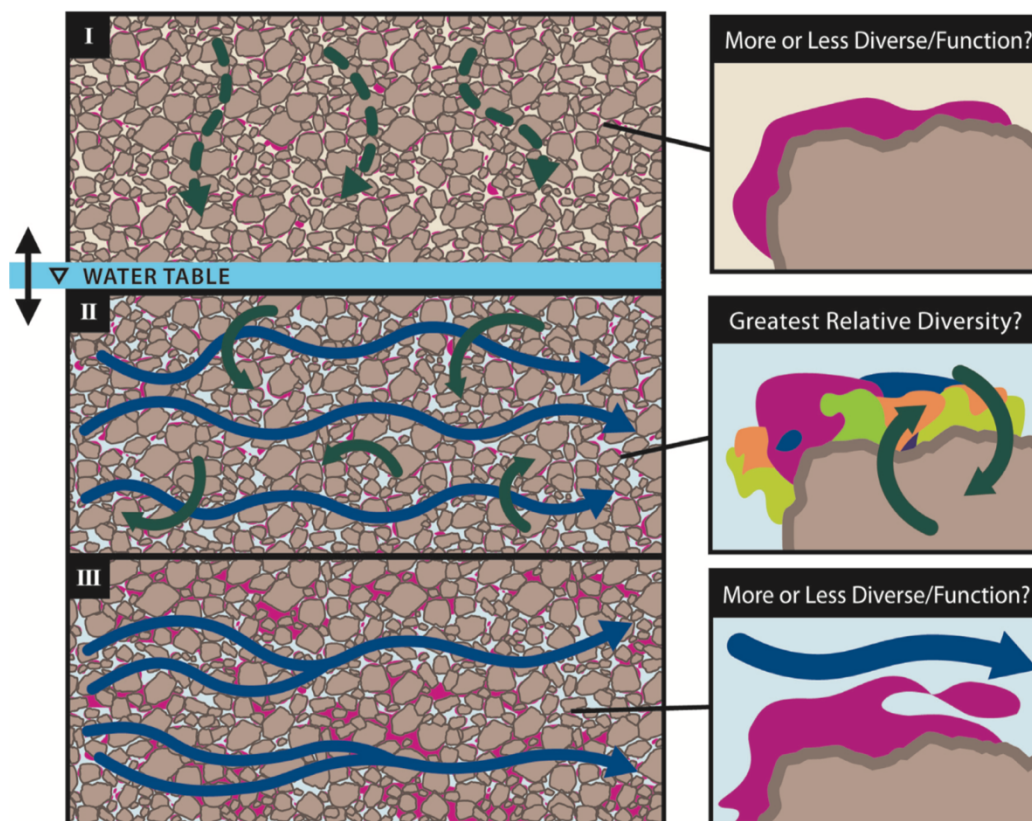


Figure E.2. Conceptual model of subsurface flow and mixing zones and potential effects on biofilm life-cycle dynamics. Subsurface porous media habitats can be conceptually divided into three zones (I, II, and III) with respect to groundwater flow and mixing. (I) The vadose zone (including the capillary fringe) is variably saturated depending on infiltration episodes and degree of vertical water table fluctuation, (II) Zone II is the “shallow” groundwater zone wherein groundwater flow, together with seasonal changes in water table elevation, can cause multi-directional flow (*i.e.* vertical and horizontal fluctuations) that can result in greater mixing, (III) the “deeper” groundwater zone (zone III) lies below the depth affected by seasonal water table fluctuations. The degree of mixing in zone III is related mainly to the ground water flow field. In zone II the higher level of seasonal mixing could result in a “hot spot” of greater relative biofilm diversity and activity (represented by multi-color sections; biofilm not depicted at scale) (Bougon et al., 2012). Zone I could have lower biofilm diversity/activity due to limited and transient mixing, although it is possible that diversity and activity in zone I would be more similar to zone II than region III. In zone III, which is deeper and has a more consistent ground water flow regime, biofilms would be less diverse/active. The roles of adhesion/detachment/dispersion could vary with the extent of mixing in the different zones and suggests that different mechanisms of microbial community assembly and diversification impact in situ biofilms.

are associated with sediment compared to planktonic cells (Hazen *et al.* 1991; Alfreider, Kroßbacher and Psenner 1997). However, the contribution of free-living and biofilm cells to subsurface processes on a per cell basis is unclear. Recently, it has been proposed that microbial competition selects against rapid growth in biofilm populations (Coyte *et al.* 2016). These findings offer a unique and contradictory perspective as to the role of free-living organisms compared to biofilms that may alter our current understanding of colonization, maintenance and dispersal of microbial populations in porous environments.

Activity in Groundwater

Most researchers have now concluded that attached bacteria dominate oligotrophic subsurface environments in terms of biomass and activity and that most planktonic cells are ‘inactive’ subsets of benthic organisms (Goldscheider, Hunkeler and Rossi 2006 and references therein). Initially, indications that groundwater samples had a low proportion of active cells came from microscopic evaluation of pristine aquifers which observed cells between 0.4 and 0.9 μm in size, suggesting that these bacteria were in a starved state with reduced activity (Balkwill and Ghiorse 1985). However, a recent study identified novel ultra-microbacteria that are inherently small ($<0.1 \mu\text{m}$) in groundwater but activity was not reported (Luef *et al.* 2015). Groundwater habitats have been shown to be able to vary drastically over time and space. For example, in a two-year study, all tested extracellular enzyme assays were found to vary significantly both spatially and temporally (Velasco-Ayuso *et al.* 2011).

Recent advances are moving away from relying solely on bulk activity measurements. Quantitative studies that are capable of linking individual microorganisms

to biogeochemical processes have been applied to groundwater from carbonate-rock aquifers. Although these results are not from a porous aquifer, the combination of metabolic labeling (*i.e.* D₂O) with Raman microspectroscopy, metaproteomics and carbon amendments quantitatively showed that naturally occurring heterotrophic organisms preferentially assimilated lignin derivatives over biomass degradation products (Taubert *et al.* 2017) and are therefore involved in subsurface carbon cycling processes.

Activity in Sediment

The overwhelming density of sediment associated organisms presents a compelling case that sediment core samples are likely the most representative samples for biomass analysis in the shallow subsurface. Studies based on cored samples have looked at the microbial activity of attached communities as a function of depth and particle size. Not only are cell numbers higher in shallower depths compared to deeper depths, the same holds true for activity (Beloin, Sinclair and Ghiorse 1988; Martino *et al.* 1998). This trend has been shown, regardless of the methodology used [ATP assays, MPN, viable plate counts and the tetrazolium reduction method (INT)]; however, activity (specifically within the saturated zone) was shown to vary seasonally dependent on the method utilized (Beloin, Sinclair and Ghiorse 1988). When comparing similar depth profiles (<50 m), other researchers have observed only slight variations in total cell abundances over depth and the largest differences were observed in the active fraction (determined with viable plate counts) which decreased with depth (Balkwill and Ghiorse 1985; Balkwill 1989). Studies utilizing radioisotope tracers have found higher metabolic activities in shallower depths as well as spikes in activity within the saturated zone (Phelps *et al.* 1988). Interestingly,

anaerobic bacteria have also been found to decrease in viability with depth and have been reported to be a 100-fold lower than aerobic organisms (Balkwill and Ghiorse 1985). Conversely, in low conductivity ecosystems, studies have found that anaerobic microorganisms have greater viability at deeper depths (Martino *et al.* 1998). The discrepancies between studies are likely attributed to differences in hydraulic conductivity which directly impacts microbial and nutrient sources and local geochemistry, the exclusion of temporal analysis and differences in methodologies.

Activity measurements comparing biofilm and planktonic populations within sediment mesocosms observed a higher proportion of activity in sediments (0.25 m columns with shallow sediments) compared to the planktonic communities (Longnecker and Kujawinski 2013), a finding that corroborates results from field studies (Thomas, Lee and Ward 1987; Hazen *et al.* 1991; Holm *etal.* 1992; Alfreider, Kroßbacher and Psenner 1997; Anneser *et al.* 2010). Some studies have observed total cellular abundances to be highest with coarse particles (Albrechtsen 1994), while others have demonstrated a greater number of microorganisms associated with fine silt particles ($<20\ \mu\text{m}$) (Harvey, Smith and George 1984). However, the study that found higher total densities of organisms associated with coarse particles showed with multiple methodologies that the greatest proportion of active organisms (91.9–100% of viable bacteria) were associated with smaller size particles (1.2–100 μm) (Albrechtsen 1994). An additional attribute that is largely unknown for sediment biofilms is cell density per given surface area and/or cooccurrence of cells or populations in more oligotrophic environments.

Paired Studies Comparing Activity in Groundwater to Sediment

In order to accurately determine the contribution of free-living and attached, paired groundwater and core samples are necessary albeit these studies are significantly fewer in number. Multiple methodologies have shown that total measured activity (*e.g.* general metabolic activity, degradation of a specific compound) is greater per gram of sediment than for comparable adjacent groundwater (per mL or L) for both contaminated and pristine aquifers (Thomas, Lee and Ward 1987; Hazen *et al.* 1991; Holm *et al.* 1992; Alfreider, Kroßbacher and Psenner 1997; Anneser *et al.* 2010). A variety of radioisotope tracers are routinely used to infer activity from environmental samples, and a study on a sandy aquifer that used ^3H -thymidine in combination with ^{14}C - leucine was not able to detect significant activity in any tested groundwater samples while activity was readily quantified for colonized material from down-well sampling devices (Alfreider, Kroßbacher and Psenner 1997). The greatest activity was observed for smaller grain size particles (125–250 μm) compared to larger particle sizes (250–500 μm) (Alfreider, Kroßbacher and Psenner 1997). The percentage of respiring bacteria as determined with INT has been shown to range between 1.0–24.9% of total cells for pumped groundwater and 6.0–41.1% of total cells for sandy sediments with on average three-fold fewer active cells in groundwater than associated with sediment. While this study demonstrated higher totals and higher numbers of metabolically active cells as well as greater rates of general metabolic activity in sediment samples, the results potentially stem from an artifact due to the use of sterile sediments suspended down-well rather than the use of actual sediment cores.

An early study used direct counts (cellular DNA stained via acridine orange) and viable plate counts to compare subsurface sediment to adjacent groundwater. Total and viable bacterial abundances for sediments collected from three individual boreholes were higher than for any adjacent groundwater samples from discrete depths (Hazen *et al.* 1991). The study also observed higher densities of total and active cells associated with aquifer sediments than groundwater and concluded that attached bacterial communities are not reflected in groundwater samples. In addition, recovered sediment isolates were capable of utilizing a broader range of carbon sources than planktonic cells from groundwater which the authors concluded relevant for *in situ* bioremediation efforts.

More recently, from a tar oil contaminated aquifer, sediments contained greater than 97.7% of all bacterial cells and displayed six-fold greater enzyme activities than groundwater, although groundwater dominated more specific processes such as sulfate and iron reduction (Anneser *et al.* 2010). These observations corroborate previous studies: the highest density of organisms is associated with sediments but groundwater communities have the potential to dominate certain redox reactions when sampled at the appropriate resolution. The idea that some microorganisms reside mainly in the planktonic phase of subsurface porous environments is supported by the predominance of methanogenic microorganisms observed in the planktonic phase compared to sediments (Lehman 2007b). Using a combination of laboratory and field-based studies Holm *et al.* (1992) examined the role of planktonic and biofilm associated cells on the biodegradation of hydrocarbons and concluded that whereas there were substantially lower rates of degradation for groundwater samples, the planktonic phase significantly contributed to the biodegradation of organic

contaminants. While reaction rates within the groundwater are typically lower on a per volume basis, the studies demonstrate the importance of sampling and studying both groundwater and subsurface biofilms.

A possible explanation for differences in activities between attached and free-living populations is likely due to differences in cell abundances. However, it remains unresolved whether free-living cells in porous subsurface habitats are in fact metabolically slower or faster. Recently, models have predicted that attached cells can be selected to grow at slower rates as to avoid mass transport limitations (Coyte *et al.* 2016). While this concept has not been directly shown under natural conditions, a study that combined cellular abundances and volumetric rates of degradation found that some cell specific activities for planktonic bacteria may be higher than for sediment associated organisms (Lehman 2007a).

Subsurface Biofilms

Subsurface communities are traditionally discussed in terms of the planktonic or attached phases with little reference to attached communities as ‘biofilms’; therefore, the following section aims to merge the available information from biofilm studies (often done in laboratory settings with single model organisms) with properties and constraints relevant to subsurface processes. Within these attached communities, microorganisms with varied metabolic functionalities can coexist and have cell- to-cell contact (Stewart and Franklin 2008), periodically detaching to become part of the planktonic phase (McDougald *et al.* 2012). It is likely these cells colonize new environments and are a primary mechanism for translocation from one surface to another (Watnick and Kolter 2000). Thus, the solid sediment matrix potentially could act as a seed bank of pelagic bacteria that can then be

translocated. Despite the ecological significance of biofilms, the relationships between source diversity (and activity) within the groundwater and local diversity (and activity) in the sediments are difficult to ascertain based upon logistics of sampling intact sediment material. Major logistics include expensive sampling that only provides a single time and space point that cannot be replicated and the subsequent impact on water flow through the disturbed matrix (discussed above).

Life Cycle Stages of Subsurface Biofilms

Biofilms appear to be an inherent phenotype for most microorganisms studied to date. The basic cycle of at least bacterial biofilms is attachment or adhesion, maturation and detachment (Hall-Stoodley, Costerton and Stoodley 2004). Typically, surface properties control cellular attachment while mass transport of substrates (influx) and/or products (efflux) limits overall biofilm maturation/growth. Detachment can be caused by a variety of conditions that result in desorption, detachment and/or dispersion dependent upon varying geochemical and geophysical conditions. It is likely that biofilms in the oligotrophic subsurface are non-continuous or patchy and that surface substrate and community composition impact adhesion. Biofilms from injections of radiolabeled cells into intact sediment cores dominated by quartz showed cell attachment around particular mineral grains suggesting a mineral preference for the adsorbed bacteria (Dong *et al.* 1999). Bacterial adhesion has been shown in laboratory studies to increase in areas where quartz sands have been artificially coated with metal oxyhydroxides (Scholl *et al.* 1990; Mills *et al.* 1994). Other studies have demonstrated different populations are enriched on different materials (Reardon *et al.* 2004; Bollmann *et al.* 2010; Converse *et al.* 2015). Further work

is needed to better understand the physical forces under low fluidization that can promote or deter microbial biofilms under oligotrophic conditions and different mixing regimes (Fig. 2).

Adhesion

Typically, microbial attachment is reversible (Dowd, Herman and Maier 2000), which could be beneficial for subsurface microorganisms subjected to environmental perturbations. The exact conditions that promote some microorganisms in the shallow subsurface to initiate attachment is unknown, but most likely includes physical (*e.g.* cell charge and flow) and chemical (*e.g.* pH and conductivity) parameters as well as biological (*e.g.* aggregation). While this has not been explicitly explored in shallow, subsurface biofilms, it is likely that the attachment of biofilms within the subsurface enhances survival in nutrient limited conditions by creating a microenvironment distinct from surrounding conditions (Coombs *et al.* 2010). Previous work mostly with *Pseudomonas* but also others (*e.g.* *Shewanella*), has compared vertical and horizontal attachment in reference to flow forces and the ability of cells to attach or detach (Conrad *et al.* 2011; Bennett *et al.* 2016). Initial attachment and any subsequent cell division has a direct impact on the architecture of the biofilm. For biofilms on porous media similar to shallow subsurface sediments and for those with slow flow regimes that impact mass transport variables (*e.g.* advection, dispersion and mass flux), the studies are quite limited.

Growth/Dispersal.

The formation and growth of biofilms in subsurface habitats is likely dependent on the microbial assemblages present, nutrient availability, substrate composition and hydraulic residence time (Coombs *et al.* 2010). Biofilm thickness is highly variable and ranges from the thickness of single cells to thicker microcolonies that are adhered together by extracellular polymeric substances. Thin and/or patchy biofilms are usually not limited by diffusion (Rittman 1993) and most likely represent biofilms in undisturbed, oligotrophic (*i.e.* pristine) subsurface environments. However, sediment biofilms can be thicker under different conditions that change over time and space, for instance, during biostimulation when nutrients are added.

Total cell numbers of bacteria that have been documented in groundwater ecosystems have ranged between 10^2 and 10^6 cells/mL of groundwater (Griebler and Lueders 2009); however, these values may be dependent on hydrological fluctuation over time (Velasco-Ayuso *et al.* 2009). While micro-eukaryotes have also been found within subsurface groundwater environments, reported observations have shown that the majority of cellular biomass in the subsurface is bacterial and archaeal (Griebler and Lueders 2009; Valster *et al.* 2009; Zinger, Gobeta and Pommiers 2012).

In comparison for subsurface matrix material, cell counts range from 10^4 to 10^9 cells/g sediment (Turco and Sadowsky 1995; Balkwill and Boone 1997; Griebler and Lueders 2009). These ranges typically vary with depth (Lin *et al.* 2012a), pH (Fierer and Jackson 2006), soil and sediment texture and porosity (Schwoerbel 1961; Balkwill and Ghiorse 1985; Strayer 1994; Hahn 2006), redox conditions, dissolved oxygen, mineral

content and moisture content (Sirisena *et al.* 2014). More recently, the distribution of viruses within subsurface groundwaters and the impact on microbial abundance has been studied (Pan *et al.* 2017). Thorough reviews on cell count abundances within different areas of the subsurface have been synthesized (Goldscheider, Hunkeler and Rossi 2006; Akob and Kussel 2011) and cell count surveys have consistently shown differences in the abundances between attached and free-living phases (Griebler and Lueders 2009) dependent upon variations in physicochemical parameters (Sinclair and Ghiorse 1989). For example, microbial population density estimates correlated positively with sand content and pore-water pH and declined with clay content and pore-water heavy metals (Sinclair and Ghiorse 1989).

In laboratory experiments, biofilm growth within granular/porous reactor systems has been shown to reduce pore spaces that leads to the blockage of pores and flow (Taylor and Jaffe' 1990a,b; Cunningham *et al.* 1991), alteration of water retention (Or *et al.* 2007) and significantly reduces hydraulic conductivity (Rodriguez-Escales *et al.* 2016). Biofilms can further reduce permeability by the entrapment of fine grained or colloidal materials that block flow (Hama 1997; Hama *et al.* 2001). It has also been shown that fine textured materials have higher occurrences of clogging compared to coarse textured materials (Vandevivere *et al.* 1995). Undoubtedly, biofilms have significant impacts on the porosity and permeability in natural porous aquifer systems; however, it is likely that biofilm heterogeneity and distribution *in situ* will be different than observed in reactor and consolidated aquifer studies. Environmental biofilms *in situ* have been shown to be patchy rather than uniform in distribution and thickness, and conceptual models have been applied

to microbial growth and transport in subsurface habitats (Vandevivere *et al.* 1995; Clement, Hooker and Skeen 1996; Ebigbo *et al.* 2010). Historically, different models have been used to estimate unsaturated (Farthing and Ogden 2017) and saturated (Molnar *et al.* 2015) water flow in porous media relevant to the shallow subsurface; however, the lack of data and understanding of microbial processes in the shallow subsurface challenges the incorporation of microbial ecology and physiology into these models. Hopefully in the future, drivers of microbial biofilm assembly and maintenance (*e.g.* selection, dispersal and drift) can be investigated and modeled with respect to hydrological parameters (*e.g.* porosity, permeability and mixing).

Laboratory Approaches to Study Subsurface Biofilms

While not a focus of this mini-review, there are numerous examples of using laboratory experiments to study subsurface microbial transport under unsaturated and saturated conditions (Tufenkji, Redman and Elimelech 2003; Jordan *et al.* 2004; Gargiulo *et al.* 2007; Harvey, Harms and Landkamer 2007; Bradford, Schijven and Harter 2015). Laboratory experiments are routinely used to mimic and investigate environmental subsurface processes (*i.e.* grain size distribution, biofilm thickness/diffusion, biodegradation, pore clogging, flow, mass transport and hydraulic conductivity), and different reactor and incubation conditions (*e.g.* column, flat plate and serum bottles) and surrogate sediments (*e.g.* silica beads/sand or collected sediment core material) are used in various combinations depending on the process being investigated. When sediment cores are taken for laboratory studies, the pore structure may be altered by packing and repacking that results in porous media flow and transport properties significantly different from *in*

situ conditions. Thus, an iterative approach combining field and laboratory studies is beneficial for ensuring laboratory findings that hold relevance to the natural system while maintaining controlled laboratory conditions necessary for developing and testing predictive models. For example, with column reactors filled with different-sized silica beads (coarse and fine), coarse sediments had higher biofilm biomass and activity although overall functionality was impaired (activity and diversity) (Perujo *et al.* 2017). Fine beads constrained biofilm activity and biomass while bead-size transitions promoted increased OM degradation and biomass at the interface (Perujo *et al.* 2017). The results corroborate the notion that particle size impacts interstitial fluxes and mixing, and thereby biofilm growth and activity for sandy sediment. Future work is needed to further elucidate these relationships under various conditions of flow, substrate flux and biofilm accumulation/activity. The hydrological impacts in the capillary fringe and water table boundary could affect biofilm dynamics in different ways that result in varying levels of biofilm diversity and activity (Fig. 2).

Hydrogeochemical mixing

As stated above, mixing can impact taxonomic diversity in groundwater communities and recent studies suggest it also affects sediment associated communities in similar ways (Ebrahimi and Or 2016, 2018). However, due to chemical (mineralogy) and physical (*e.g.* size, arrangement) heterogeneity of the sediment matrix, niche partitioning and species filtering are likely additional factors that impact the composition of attached communities. In addition, mixing in the shallow subsurface due to faster and shorter local flow paths could impact local hydrodynamic dispersion and thus biofilms. Therefore,

sediment biofilms likely have distinct zone-specific responses (*e.g.* vadose to capillary fringe to saturated zones) (Fig. 1). Also, the subsurface sediment zones likely experience different degrees and rates of flow that impact the formation and stability (chemical, physical, and biological) of sediment-associated biofilms, and the biofilms are impacted by fluctuations of the water table and associated redistribution in the capillary fringe (Moser *et al.* 2003; Stegen *et al.* 2016).

Based on the available information from diversity-based studies, bulk activities and biofilm studies, hydrodynamics likely affects biofilm structure, function and dispersal in shallow subsurface aquifers both by vertical and horizontal mixing. In high shear environments (*e.g.* water distribution lines), high shear stress is observed to decrease biofilm diversity and thickness (Rochex *et al.* 2008); however, the relationship between mixing and biofilm diversity is not known for low-shear conditions analogous to shallow subsurface environments that can have unsteady groundwater flow (Sposito 2006). If required resource ratios are not available, microbial activity cannot be sustained and continued non-growth could promote dispersal and/or death. Some work has attempted to explain the occurrence of microbial populations in terms of a resource ratio theory, where a given level of resource is needed to sustain a population (*i.e.* a consumption rate that is greater than a death rate at a given resource concentration) (Smith 1993; De Mazancourt and Schwartz 2010). However, this relationship does not account for varying substrate affinities, interacting populations or different behavior across phase boundaries under mixing conditions. Recently, it has been hypothesized that biofilm cells with restricted growth can outcompete populations with faster growth in the bulk-phase based upon a

laboratory model (Coyte *et al.* 2016). This is an interesting hypothesis to test relevant conditions for the shallow subsurface that includes different populations, interactions and/or activities in a porous medium with dynamic mixing in the shallow and deep saturated zone (Fig. 2).

Microbial Interactions Within Biofilms

Many shallow subsurface biofilms are likely comprised of multiple species as in numerous other environments; however, few studies have delineated the spatial arrangement of microbial cells on particles from the shallow subsurface nor have deciphered potential metabolic interactions *in situ*. Certainly, laboratory studies of multispecies biofilms observe that populations are not always randomly distributed but organized based on needs (Møller *et al.* 1998; Watnick and Kolter 2000). Conducting such studies with native material under *in situ* conditions are challenging. Recent work with upper layer soil/sediment particles have shown particle-specific communities (Jackson and Weeks 2008; Hemkemeyer *et al.* 2015, 2018), but similar work for shallow subsurface sediments (>1 m depth) is sparse. From cell counts, one cannot determine whether attached populations reside as individual cells separated by micrometers of space, as clonal microcolonies or as multispecies biofilms over preferred locations (*e.g.* nutrient/mineral availability). This is also an issue with community analyses via amplicon or shotgun sequencing. Samples large enough to yield sufficient DNA quantities often encompass too much physical space to confidently infer representative microscale interactions, although progress has been made with upper layer soil particles. Future work is needed to elucidate whether sediments from the shallow subsurface are amenable to the same methods and if

microscopy methods can be used with intact sediment samples in order to retain inherent physical structure.

Future Directions

It has become increasingly apparent that free-living and biofilm associated cells have distinct physiologies and function but the potential impacts on shallow subsurface systems is not well understood (Hall-Stoodley, Costerton and Stoodley 2004; Anneser *et al.* 2010). Many questions remain regarding the biofilm 'life-cycle' including attachment transitions, the distribution and rate of specialized and general activity, cooperative/competitive interactions, and mechanisms of dispersal in the shallow subsurface mixing zones (Fig. 2). Due to sampling challenges and the complexity of the heterogeneous subsurface matrix that ranges across the vadose, capillary fringe and saturated zones, few field sites have been comprehensively described and studied despite the important ecosystem services associated with shallow subsurface systems. The shallow subsurface has historically been considered a stable environment, but it is now clear that temporal and seasonal dynamics influence hydrological mixing, particularly between and within the saturated and capillary fringe zones. Aquifer recharge and fluctuating water table can occur via seasonal patterns, and not surprisingly, the transition zones between the variably saturated and saturated zones has been shown to be an important ecotone for microbial diversity and activity. Due to the complexity of the system and logistical challenges in sampling, there is much yet to be learned about the distribution of shallow subsurface biofilms, the physiological activities/responses to environmental disturbances

related to geochemical cycling and the roles these systems play in groundwater maintenance and stability. Technological advances for sample retrieval and fine-scale analyses (spatial, temporal, cellular) of subsurface samples are needed, including samplers that can retrieve intact porous media and associated biofilms. Similarly, engineered reactor systems need to be modified to accurately simulate subsurface environmental conditions and address inconsistencies in reproducibility that currently exist.

In order to understand and predict the role of microorganisms accurately within an environmental context, it is essential to distinguish between active and inactive organisms. The majority of studies on activity in subsurface porous aquifers are from the 1980–1990's and very few recent studies incorporate activity measurements with sequencing technologies. While the more recently adopted metagenomics-based sequencing approaches have opened new windows as to the functional diversity present within porous aquifers, activity is seldom linked to phylogeny. Of the open reading frames recovered with metagenomic sequencing, typically the functions from a small fraction can be linked to known genes and only a few of these genes have been studied in depth (Ferrer *et al.* 2016). Currently, untargeted metagenomic sequencing predominantly retrieves genomic sequences from dominant organisms and does not allow active organisms to be differentiated from inactive. While this is more suitable for low diversity habitats (Tyson *et al.* 2004; Woyke *et al.* 2006), subsurface environments can be highly diverse (Hug *et al.* 2015). Therefore, in order to accurately capture rare, underexplored and possibly environmentally significant metabolic processes, it will be imperative to apply functional/targeted metagenomic approaches. While the use of sequencing technologies

has allowed microbial ecologists to glean the taxonomic compositions of microbial communities as never before, it is important to note that such methodologies contain inherent problems (Wintzingerode, Goebel and Stackebrandt 1997; Bent and Forney 2008; Fraser *et al.* 2009). Recently, Carini *et al.* (2016) showed that extracellular, or 'relic', DNA was abundant in a variety of soil and sediment types and that this DNA can skew diversity measurements.

Due to the rapid evolution of technologies capable of working with small DNA and transcript quantities, the application of targeted strategies (*i.e.* sequencing data within a functional context) has increased for environmental studies (Lueders *et al.* 2016). Approaches already exist that target active fractions of microbial communities (*e.g.* bioorthogonal non-canonical amino acid tagging (BONCAT) (Hatzenpichler and Orphan 2015), DNA and RNA stable isotope probing (SIP) (Lueders *et al.* 2016) and propidium monoazide (PMA)-Seq (Carini *et al.* 2016)) which can then be combined with metagenomic or rRNA sequencing strategies. In the near future, these methods will be combined with microscopic/spectroscopic techniques that allow physical structure to be maintained and the proper scale for sediment-associated biofilms to be determined.

The ability to infer subsurface-specific functional capabilities from genetic information, as well as the generation of testable hypotheses that can be confirmed at ecologically relevant microscales, is limited by the current lack of subsurface-specific reference sequences. In addition, laboratory studies are needed at the microscale with field-relevant isolates to confirm hypotheses generated from sequencing data. Of the currently available subsurface isolates, many have slow growth rates and/or most likely use forms of

C, N and P associated with sediments not typically used in cultivation/microcosms. It is becoming increasingly crucial to frame *in situ* experimentation to ecological questions and conditions pertinent to the respective environment (*e.g.* subsurface transport through porous media with intermittent inputs of temporally- and spatially-relevant OM). As noted by Prosser (2012), many microbial ecology questions require studies that focus on smaller spatial scale, phenotypic diversity, temporality and activity/rates. Through these combined approaches, a more complete understanding of shallow subsurface ecosystems will be gained that includes biofilm dynamics at zone interfaces.

Funding

This material by ENIGMA-Ecosystems and Networks Integrated with Genes and Molecular Assemblies (<http://enigma.lbl.gov>), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231.

REFERENCES CITED

- Akob, DM., Küsel, K., 2011. Where microorganisms meet rocks in the Earth's Critical Zone. *Biogeosciences* 8, 3531-3543.
- Alexander, M., Introduction to Soil Microbiology. John Wiley and Sons, New York 1977.
- Alm Laboratories. SmileTrain. <https://github.com/almlab/SmileTrain/wiki/>. Accessed 1 July 2016.
- Amann, R. and Rosselló-Móra, R., 2016. After all, only millions?. *MBio*, 7(4), pp.e00999-16.
- Anantharam K, Brown CT, Hug LA et al., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communication*. 7:13219. DOI: 10.1038/ncomms13219
- Anderson, RT., Lovley, DR., 1997. Ecology and biogeochemistry of in situ groundwater bioremediation. *Advances in Microbial Ecology* 15, 289-350.
- Andersen, T., Carstensen, J., Hernandez-Garcia, E. and Duarte, C.M., 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution*, 24(1), 49-57.
- Atekwana, Estella A., D. Dale Werkema, and Eliot A. Atekwana., 2006. "Biogeophysics: The effects of microbial processes on geophysical properties of the shallow subsurface." In *Applied hydrogeophysics*, pp. 161-193. Springer, Dordrecht.
- Baldock, D., Nebe-von-Caron, G., Bongaerts, R. and Nocker, A., 2013. Effect of acidic pH on flow cytometric detection of bacteria stained with SYBR Green I and their distinction from background. *Methods and applications in fluorescence*, 1(4), p.045001.
- Balkwill DL, Boone DR., 1997. Identity and diversity of microorganisms cultured from subsurface environments. In: Amy PS, Haldeman DL (eds.). *The Microbiology of the Terrestrial Deep Subsurface*. Boca Raton: CRC Press, 105-117.
- Bates, DM., 2010. *lme4: Mixed-effects modeling with R*. Springer. <http://lme4.r-forge.r-project.org/book/front.pdf>
- Bell, T., Newman, J.A., Silverman, B.W., Turner, S.L. and Lilley, A.K., 2005. The contribution of species richness and composition to bacterial services. *Nature*, 436(7054): pp. 1157.

- Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F., 2012. Impacts of climate change on the future of biodiversity. *Ecology Letters* 15: 365-377.
- Beman JM, Steele JA, Fuhrman JA., 2011. Co-occurrence patterns for abundant marine archaeal and bacterial lineages in the deep chlorophyll maximum of coastal California. *ISME J* 5: 1077–1085.
- Ben Maamar S, Aquilina L, Quaiser A, Pauwels H, Michon-Coudouel S, Vergnaud-Ayraud V, Labasque T, Roques C, Abbot BW, Dufresne A., 2015. Groundwater isolation governs chemistry and microbial community structure along hydrologic flowpaths. *Frontiers in Micro* 6: 1-13
- Bence, J.R., 1995. Analysis of short time series: correcting for autocorrelation. *Ecology*, 76(2), pp.628-639.
- Bent, S.J., Forney, L.J., 2008. The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J* 2 (7), 689–695.
- Berkowitz B, Silliman SE, Dunn AM., 2004. Impact of the capillary fringe on local flow, chemical migration, and microbiology. *Vadose Zone J*; 3:534-48.
- Berkowitz, B., Dror, I. and Yaron, B., 2014. Characterization of the Subsurface Environment. In *Contaminant Geochemistry*, pp. 3-28. Springer, Berlin, Heidelberg.
- Bernard, G., Pathmanathan, J.S., Lannes, R., Lopez, P. and Baptiste, E., 2018. Microbial dark matter investigations: how microbial studies transform biological knowledge and empirically sketch a logic of scientific discovery. *Genome biology and evolution*, 10(3), pp.707-715.
- Bienhold, C., Boetius, A., Ramette, A., 2011. The energy-diversity relationship of complex bacterial communities in Arctic deep-sea sediments. *ISME J* 6 (4), 724-732.
- Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. 2013. Evaluationg rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J*. 7(11) ppgs 2061-8
- Borcard D, Gillet F, Legendre P. 2011. *Numerical Ecology with R*. Springer
- Bougon N, Auterives C, Aquilina L, et al. 2010. Nitrate and sulphate dynamics in peat subjected to different hydrological conditions: Batch experiments and field comparisons. *Journal of Hydrology* . 411:1-2, pp. 12-24.

- Bougon, N., Aquilina L, Molenat J., et. al., 2011. Influence of depth and time on diversity of free-living microbial community in the variably saturated zone of a granitic aquifer. *FEMS Microbiol Ecol* 80, 98-113.
- Brady, NC and Weil, RR. 2014. *Elements of the nature and properties of soils*, 3rd ed. Prentice Hall, New Jersey. ISBN 978-0-13-501433-2
- Briggs, M. A., F. D. Day-Lewis, J. P. Zarnetske, and J. W. Harvey (2015), A physical explanation for the development of redox microzones in hyporheic flow, *Geophys. Res. Lett.*, 42, 4402–4410, doi:10.1002/2015GL064200.
- Brown CT, Hug LA, Thomas BC et al., 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*; 523:208-211.
- Campbell KM, Kukkadapu RK, Qafoku NP et al. Geochemical, mineralogical and microbiological characteristics of sediment from a naturally reduced zone in a uranium-contaminated aquifer. *Appl Geochem* 2012;27: 1499-1511.
- Caporaso, JG., Lauber, CL., Walters, WA., et al., 2010. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS* 108, 4516-4522.
- Caporaso, JG., Lauber, CL., Walters, WA., et al., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6 (8), 1621-1624.
- Carini P, Marsden PJ, Lef JW et al. 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiol* 2016;2:16242, DOI: 10.1038/nmicrobiol.2016.242.
- Castelle CJ, Hug LA, Wrighton KC et al., 2013. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. *Nat Commun*; 4:2120, DOI: 10.1038/ncomms3120.
- Chakraborty, R., Wu, C.H. and Hazen, T.C., 2012. Systems biology approach to bioremediation. *Current Opinion in Biotechnology*, 23(3), pp.483-490.
- Chapelle, FH., 2000. The significance of microbial processes in hydrogeology and geochemistry, *Hydrogeology Journal* 8(1), 41–46.
- Chase JM, Myers JA. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. *Philos Trans R Soc Lond B Biol Sci* 366:2351–2363. <https://doi.org/10.1098/rstb.2011.0063>.

- Chase, JM., Kraft, NJB., Smith, KG., et al., 2011. Using null models to disentangle variation in community dissimilarity from variation in alpha-diversity. *Ecosphere* 2, 1-11.
- Chen C., Hemme C., Beleno J., Shi Z. J., Ning D., Qin Y., Tu Q., Jorgensen M., He Z., Wu L., and Zhou J. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. *The ISME Journal* 12, 1210-1224 (2018) doi:10.1038/s41396-017-0037-1.
- Chesson P. 2000. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* 31:343-366. <https://doi.org/10.1146/annurev.ecolsys.31.1.343>.
- Chu H, Sun H, Tripathi BM et al., 2016. Bacterial community dissimilarity between the surface and subsurface soils equals horizontal differences over several kilometers in the western Tibetan Plateau. *Environ Microbiol*; 18:1523-33.
- Clarke, JA, Mirza, S., 2006. A comparison of some common methods for detecting Granger noncausality. *Journal of Statistical Computation and Simulation* 76 (3), 207-231. doi: 10.1080/10629360500107741.
- Cottrell, M. T., & Kirchman, D. L. 2000. Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low-and high-molecular-weight dissolved organic matter. *Appl. Environ. Microbiol.*, 66(4), 1692-1697.
- Coyte KZ, Tabuteau H, Gaffney EA et al., 2016. Microbial competition in porous environments can select against rapid biofilm growth. *Proc Natl Acad Sci USA*; 114:E161–E170.
- Crawford, J.W., Deacon, L., Grinev, D., Harris, J.A., Ritz, K., Singh, B.K. and Young, I., 2011. Microbial diversity affects self-organization of the soil–microbe system with consequences for function. *Journal of the Royal Society interface*, 9(71), pp.1302-1310.
- Crump BC, Amaral-Zettler LA, Kling GW. 2012. Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J* 6: 1629-1639
- Crump, BC., Peterson, BJ., Raymond, PA, Amon RMW, Rinehart A, McClelland JW., et al., 2009. Circumpolar synchrony in big river bacterioplankton. *Proc Natl Acad Sci USA* 106:, 21208–21212.

- Danielopol, D.L., Griebler, C., Gunatilaka, A., Notenboom, J., 2003. Present state and future prospects for groundwater ecosystems. *Environmental Conservation* 30,104–130.
- Davinic, M., Fultz, L.M., Acosta-Martinez, V., Calderón, F.J., Cox, S.B., Dowd, S.E., Allen, V.G., Zak, J.C. and Moore-Kucera, J., 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. *Soil Biology and Biochemistry*, 46, pp.63-72.
- Dieser M, Nocker A, Priscu JC, Foreman CM. 2010. Viable microbes in ice: application of molecular assays to McMurdo Dry Valley lake ice communities. *Antarctic Science* 22(5), pgs 470-476.
- Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen, L.G. and Loescher, H.W., 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*, 115(7), pp.1424-1432.
- Ebrahimi A, Or D. 2016. Microbial community dynamics in soil aggregates shape biogeochemical gas fluxes from soil profiles-upscaling an aggregate biophysical model. *Glob Chang Biol*;22:3141-56.
- Ebrahimi A, Or D. 2018. On upscaling of microbial processes and biogeochemical fluxes from aggregates to landscapes. *J Geophys Res Biogeosci*;12:1526-47.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200
- Edgar RC., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* 10, 996-998.
- Eiler, A., Heinrich, F., Bertilsson, S. 2012. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J* 6: 330–342.
- Environmental Protection Agency, website. Accessed March 4, 2019.
<https://cumulis.epa.gov/supercpad/SiteProfiles/index.cfm?fuseaction=second.cleanup&id=0404152>
- Farjalla VF, Srivastava DS, Marino NA, et al. 2012. Ecological determinism increases with organism size. *Ecology* 93:7, pp 1752-1759

- Faust, K., Lahti, L., Gonze, D., De Vos, W.M. and Raes, J., 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Current opinion in microbiology*, 25, pp.56-66.
- Field, S.A., O'Conner, P.J., Tyre, A.J. and Possingham, H.P., 2007. Making monitoring meaningful. *Austral Ecology*, 32(5), pp.485-491.
- Field, J.P., Breshears, D.D., Law, D.J., Villegas, J.C., López-Hoffman, L., Brooks, P.D., Chorover, J., Barron-Gafford, G.A., Gallery, R.E., Litvak, M.E. and Lybrand, R.A., 2015. Critical Zone services: Expanding context, constraints, and currency beyond ecosystem services. *Vadose Zone Journal*, 14(1).
- Fields, MW., Bagwell, CE., Carroll, SL., et al., 2006. Phylogenetic and functional biomarkers as indicators of bacterial community responses to mixed-waste contamination. *Environ Sci Technol* 40, 2601–2607.
- Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103: 626-631
- Fierer, N., 2008. Microbial biogeography: patterns in microbial diversity across space and time. In *Accessing uncultivated microorganisms* (pp. 95-115). American Society of Microbiology.
- Fine, PVA., Kembel, SW., 2011. Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. *Ecography* 34, 552-565.
- Fittipaldi M, Nocker A, Codony F. 2012. Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. *Journal of Microbiological Methods*, 91, pp. 276-289.
- Flynn, TM., Sanford, RA., Ryu, H., 2013. Functional microbial diversity explains groundwater chemistry in a pristine aquifer. *BMC Microbiology* 13, 146.
- Forsythe, GE., Malcolm, MA., Moler, CB., 1977. *Computer Methods for Mathematical Computations*. Prentice-Hall Professional Technical Reference.
- Franklin, R.B., Garland, J.L., Bolster, C.H. and Mills, A.L., 2001. Impact of dilution on microbial community structure and functional potential: comparison of numerical simulations and batch culture experiments. *Appl. Environ. Microbiol.*, 67(2), pp.702-712.

- Freckman, D.W., Blackburn, T.H., Brussaard, L., Hutchings, P., Palmer, M.A. and Snelgrove, P.V., 1997. Linking biodiversity and ecosystem functioning of soils and sediments. *Ambio*, pp.556-562.
- Fredrickson, J. K., et al. 1997, Pore-size constraints on the activity and survival of subsurface bacteria in a late Cretaceous shale-sandstone sequence, northwestern New Mexico, *Geomicrobiol. J.*, 14(3), 183–202.
- Fredrickson, J.K. and Fletcher, M., 2001. Subsurface microbiology and biogeochemistry. Wiley-Liss.
- Freedman Z, Zak DR. 2015 Soil bacterial communities are shaped by temporal and environmental filtering: evidence from a long-term chronosequence. *Environmental Microbiology* 17:9, pgs. 3208-3218.
- Friedman J, Alm EJ 2012 Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput Biol* 8(9): e1002687. doi:10.1371/journal.pcbi.1002687
- Fuhrman JA, Cram JA, Needham DM: Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* 2015, 13:133-146.
- Fuhrman, JA., Hewson, I., Schwalbach, MS, Steele JA, Brown MV, Naeem S., 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* 103:, 13104–13109 .
- Fuhrman, JA., 2009. Microbial community structure and its functional implications. *Nature* 459:, 193–199. <http://dx.doi.org/10.1038/nature08058>.
- Ganoulis, J. and Fried, J., 2018. Transboundary Hydro-Governance: From Conflict to Shared Management. Springer.
- Galand PE, Casamayor EO, Kirchman DL, Lovejoy C., 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci USA* 106: 22427–22432.
- Gerth, J., Forstner, U., 2004. Part IV: Fitness for aquatic systems long-term forecast: Key to groundwater protection. *Environ Sci & Pollut Res* 11, 49-56.
- Gilbert JA, Steele JA, Caporaso JG, et. al., 2011. Defining seasonal marine microbial community dynamics. *ISME J*, doi: 10.1038/ismej.2011.107.
- Gleick PH, et al., 2018. The World's Water: The Report on Freshwater Resources (Pacific Institute, Oakland, CA), Vol 9.

- Gniese, C., Bombach, P., Rakoczy, J., Hoth, N., Schlömann, M., Richnow, H.H. and Krüger, M., 2014. Relevance of deep-subsurface microbiology for underground gas storage and geothermal energy production. In *Geobiotechnology II* (pp. 95-121). Springer, Berlin, Heidelberg.
- Goldscheider, N., Hunkeler, D., Rossi, P., 2006. Microbial biocenoses in pristine aquifers and an assessment of investigation methods. *Hydrogeol. J.* 14, 926-941.
- Gonzalez, A., King, A., Robeson II, M.S., Song, S., Shade, A., Metcalf, J.L. and Knight, R., 2012. Characterizing microbial communities through space and time. *Current opinion in biotechnology*, 23(3), pp.431-436.
- Griebler, C., Avramov, M., 2015. Groundwater ecosystem services: a review. *Freshwater Science* 34, 355-367.
- Griebler, C., Lueders, T., 2009. Microbial biodiversity in groundwater ecosystems. *Freshwater Biology* 54, 649–77.
- Griffiths, B.S., Ritz, K., Wheatley, R., Kuan, H.L., Boag, B., Christensen, S., Ekelund, F., Sørensen, S.J., Muller, S. and Bloem, J., 2001. An examination of the biodiversity–ecosystem function relationship in arable soil microbial communities. *Soil Biology and Biochemistry*, 33(12-13), pp.1713-1722.
- Haack, SK., Fogarty, LR, West, TG., et al., 2004. Spatial and temporal changes in microbial community structure associated with recharge-influenced chemical gradients in a contaminated aquifer. *Environ. Microbiol.* 6, 438-448.
- Haegeman B, Hamelin J, Moriarty J et al. 2016. Robust estimation of microbial diversity in theory and in practice. *ISME J*, 7, 1092-1101. Doi:10.1038/ismej.2013.10.
- Hammer, O., Harper, D., Ryan, PD., 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4 (1),9.
- Hao YQ, Zhao XF, Zhang DY. 2015 Field experimental evidence that stochastic processes predominate in the initial assembly of bacterial communities. *Environ Microbiol* 1-10.
- Hatzenpichler, R., Connon, S., Goudeau, D, Malmstrom RR, Woyke T and Orphan V., 2016. *Proc Natl Acad Sci USA* 113:, E4069-E4078.
- Hay, M. B., D. L. Stoliker, J. A. Davis, and J. M. Zachara. 2011. Characterization of the intragranular water regime within subsurface sediments: Pore volume, surface area, and mass transfer limitations, *Water Resour. Res.*, 47, W10531, doi:10.1029/2010WR010303.

- Hazen, T.C., Jiménez, L., de Victoria, G.L. and Fliermans, C.B., 1991. Comparison of bacteria from deep subsurface sediment and adjacent groundwater. *Microbial Ecology*, 22(1), pp.293-304.
- Hazen, TC., Dubinsky EA., DeSantis TZ, et. al., 2010. Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. *Science* 330 (6001), 204-208. DOI: 10.1126/science.1195979
- Hefley, T.J., Broms, K.M., Brost, B.M., Buderman, F.E., Kay, S.L., Scharf, H.R., Tipton, J.R., Williams, P.J. and Hooten, M.B., 2017. The basis function approach for modeling autocorrelation in ecological data. *Ecology*, 98(3), pp.632-646.
- Hemme, C.L., Tu, Q., Shi, Z., Qin, Y., Gao, W., Deng, Y., Nostrand, J.D.V., Wu, L., He, Z., Chain, P.S. and Tringe, S.G., 2015. Comparative metagenomics reveals impact of contaminants on groundwater microbiomes. *Frontiers in microbiology*, 6, p.1205.
- Herrmann M, Rusznyak A, Akob DM et al. Large fractions of CO₂-fixing microorganisms in pristine limestone aquifers appear to be involved in the oxidation of reduced sulfur and nitrogen compounds. *Appl Environ Microbiol* 2015;81:2384-94.
- Hildenbrandt, ZL., Carlton, DD., Fonteno, BE, et. al. (2015).. A Comprehensive Analysis of Groundwater Quality in The Barnett Shale Region. *Environ. Sci. Technol.* 49 (13), 825-8246. DOI: 10.1021/acs.est.5b01526
- Hirsch PR, Mauchline TH, Clark IM. 2010. Culture-independent molecular techniques for soil microbial ecology. *Soil Biology and Biochemistry*. 42(6) pgs. 878-887.
- Hofle, MG., Haas, H., Dominik, K., 1999. Seasonal dynamics of bacterioplankton community structure in a eutrophic lake as determined by 5S rRNA analysis. *Appl Environ Microbiol* 65, 3164–3174.
- Holliger, C., Gaspard, S., Glod, G., Heijman, C., Schumacher, W., Schwarzenbach, R.P. and Vazquez, F., 1997. Contaminated environments in the subsurface and
- Hubalek V, Wu X, Eiler A et al. Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian Shield. *ISME J* 2016;10:2447-58.
- Hubbell SP. 2001. The unified neutral theory of biodiversity and bio- geography, vol 32. Princeton University Press, Princeton, NJ.

- Hug L, Thomas BC, Brown CT, et al (2015). Aquifer environment selects for microbial species cohorts in sediment and groundwater. *ISME J*, 9: pgs 1846-1856.
- Hughes, B.B., Beas-Luna, R., Barner, A.K., Brewitt, K., Brumbaugh, D.R., Cerny-Chipman, E.B., Close, S.L., Coblenz, K.E., De Nesnera, K.L., Drobnitch, S.T. and Figurski, J.D., 2017. Long-term studies contribute disproportionately to ecology and policy. *BioScience*, 67(3), pp.271-281.
- Humbert, JF., Dorigo, U., 2005. Biodiversity and aquatic ecosystem functioning: a mini review. *Aquatic Ecosys. Health Management* 8, 367-374.
- Hwang, C., Wu, W., Gentry, TJ., et al., 2009. Bacterial community succession during in situ uranium bioremediation: spatial similarities along controlled flow paths. *ISME J* 3, 47-64.
- Hyndman, and L.J. Forney. TG., et al., 2004. Spatial and temporal changes in microbial community structure associated with recharge-influenced chemical gradients in a contaminated aquifer. *Environ. Microbiol.* 6:, 438-448.
- Jackson, C.R. and Weeks, A.Q., 2008. Influence of particle size on bacterial community structure in aquatic sediments as revealed by 16S rRNA gene sequence analysis. *Appl. Environ. Microbiol.*, 74(16), pp.5237-5240.
- Jackson, P.M. and Smith, L.K., 2014. Exploring the undulating plateau: the future of global oil supply. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2006), p.20120491.
- Jagers, P. and Klebaner, F., 2003. Random variation and concentration effects in PCR. *Journal of Theoretical Biology*, 224(3), pp.299-304.
- Jiang, Y., Qian, H., Wang, X., Chen, L., Liu, M., Li, H. and Sun, B., 2018. Nematodes and microbial community affect the sizes and turnover rates of organic carbon pools in soil aggregates. *Soil Biology and Biochemistry*, 119, pp.22-31.
- Jones AA & Bennett PC., 2014. Mineral microniches control the diversity of subsurface microbial populations. *Geomicrobiol*; 31:246-61.
- Jones RM, Goordial JM and Orcutt BN. 2018. Low Energy Subsurface Environments as Extraterrestrial Analogs. *Front. Microbiol.* 9:1605. doi: 10.3389/fmicb.2018.01605
- Jones, S.E. and Lennon, J.T., 2010. Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences*, 107(13), pp.5881-5886.

- Jost L. 2006. Entropy and diversity. *OIKOS* 113(2): 363-375
- Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., Kusel, K., Rillig, MC., Rivett, DW., Salles, JF., van der Heijden, MGA., Youseff, NH., Zhang, X., Wei, Z., Gera Hol, WH. 2017. Where less may be more: how the rare biosphere pulls ecosystem strings. *ISME J.* 11, 853-862.
- Kaiser, J.P. and Bollag, J.M., 1990. Microbial activity in the terrestrial subsurface. *Experientia*, 46(8), pp.797-806.
- Karl DM, Church MJ, Dore JE, Letelier RM, Mahaffey C., 2012. Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proc Natl Acad Sci USA* 109: 1842–1849
- Kebschull, J.M. and Zador, A.M., 2015. Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic acids research*, 43(21), pp.e143-e143.
- Keddy, P.A., 1992. Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science*, 3(2), pp.157-164.
- Kellermann, C., Selesi, D., Lee, N., Hügler, M., Esperschütz, J., Hartmann, A., and Griebler, C. (2012) Microbial CO₂ fixation potential in a tar-oil-contaminated porous aquifer. *FEMS Microbiology Ecology* 81: 172-187
- Kesler, S.E., 2007. Mineral supply and demand into the 21st century. In *Proceedings for a Workshop on Deposit Modeling, Mineral Resource Assessment, and Their Role in Sustainable Development*. Circular (Vol. 1294, pp. 55-62).
- Kim, YJ., Moon, JW., Roh, Y., Brooks, SC., 2009. Mineralogical characterization of saprolite at the FRC background sites in Oak Ridge, Tennessee. *Environ Geol* 58:1301-1307. DOI. doi: 10.1007/s00254-008-1633-1
- King, A., Preheim SP., Bailey KL., et al., 2017. Temporal dynamics of in-field bioreactor populations reflect the groundwater system and respond predictably to perturbation. *Environ. Sci. Technol.* 51, 2879–2889
- Koenig, JE., Spor, A., Scalfone, N., et al., 2011. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci USA* 108, 4578-85.
- Kraft NJB, Adler PB, Godoy O, James EC, Fuller S, Levine JM. (2015) Community assembly, coexistence and the environmental filter metaphor. *Funct Ecol* 29: 592-599

- Langendhede S, Lindstrom ES, Tranvik LJ. 2006. Structure and function of bacterial communities emerging from different sources under identical conditions. *Applied and Environmental Microbiology*. 72:1, Pgs. 212-220.
- Langendhede S and Szekely AJ. 2011. Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J* 5: 1086-1094.
- Langwaldt, JH, Puhakka, JA., 2000. On-site biological remediation of contaminated groundwater: a review. *Environmental Pollution* 107, 187-197.
- Lauber CL, Ramirez KS, Aanderud Z et al., 2012. Temporal variability in soil microbial communities across land-use-types. *ISME J*, 7: pgs 1641-1650.
- Lauber, C.L., Hamady, M., Knight, R. and Fierer, N., 2009. Soil pH as a predictor of soil bacterial community structure at the continental scale: a pyrosequencing-based assessment. *Applied and Environmental Microbiology*.
- Lazar CS, Stoll W, Lehmann R et al. Archaeal diversity and CO₂ fixers in carbonate-/siliciclastic-rock groundwater ecosystems. *Archaea* 2017;2017:2136287.
- Lee JE, Buckley HL, Etienne RS, Lear G. 2013. Both species sorting and neutral processes drive assembly of bacterial communities in aquatic microcosms. *FEMS Microbiol Ecol* 86: pgs 288-302.
- Lee KS (2013) *Underground thermal energy storage*. Springer, London
- Lee, JH., Lee, BJ., 2017. Microbial reduction of Fe (III) and SO₄²⁻ and associated microbial communities in the alluvial aquifer groundwater and sediments. *Microbiol Ecol* 76, 182-91.
- Legendre, P. and Legendre, L.F., 2012. *Numerical Ecology* (Vol. 24). Elsevier.
- Leizeaga, A., Estrany, M., Forn, I. and Sebastián, M., 2017. Using click-chemistry for visualizing in situ changes of translational activity in planktonic marine bacteria. *Frontiers in microbiology*, 8, p.2360.
- Lennon JT, Locey KJ., 2016. The underestimation of global microbial diversity. *mBio* 2016;7:e01298-16, DOI: 10.1128/mBio.01298-16.
- Lesnoff, M., Lancelot, R., 2012. aod: Analysis of Overdispersed Data. R package version 1.3, URL <http://cran.r-project.org/package=aod>

- Levy-Booth, DJ et al., 2007. Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*. 39, 2977–2991.
- Liang, Y.T., Z. H.YT., He, L. Y.ZH., Wu, Y. Deng, G.H. Li, and J.-Z. Zhou.LY, Y., et al., 2010. Development of a Common Oligonucleotide Reference Standard for microarray data normalization and comparison across different microbial communities. *Appl. Environ. Microbiol*, 76 (4:), 1088-94.
- Liang, YT., He, ZH., Wu, LY, Y., et al., 2010. Development of a common oligonucleotide reference standard for microarray data normalization and comparison across different microbial communities. *Appl. Environ. Microbiol*, 76 (4), 1088-94.
- Lin X, Kennedy D, Peacock A et al. Distribution of microbial biomass and potential for anaerobic respiration in Hanford Site 300 Area subsurface sediment. *Appl Environ Microb* 2011;7:759–67.
- Lin XJ, Kennedy D, Fredrickson J, et al. Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environ Microbiol* 2012b;14:414-25.
- Lin, X., McKinley J., Resch CT., et al., 2012. Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. *ISME J* 6(9), 1665-1676.
- Liu, C., Liu, Y., Kerisit, S. and Zachara, J., 2015. Pore-scale process coupling and effective surface reaction rates in heterogeneous subsurface materials. *Reviews in Mineralogy and Geochemistry*, 80(1), pp.191-216.
- Locey KJ, Lennon JT. 2016 Scaling laws predict global microbial diversity. *PNAS* 113:21, pg 5970-5975
- Loreau M, et. al. 2001. Ecology - biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* 294: 804-808
- Lueders, T., Dumont, M. G., Bradford, L., and Manefield, M. (2016). RNA- stable isotope probing: from carbon flow within key microbiota to targeted transcriptomes. *Curr. Opin. Biotechnol.* 41, 83–89.
doi: 10.1016/j.copbio.2016.05.001
- Lui, G.C., Li, W.K., Leung, K.M., Lee, J.H. and Jayawardena, A.W., 2007. Modelling algal blooms using vector autoregressive model with exogenous variables and long memory filter. *Ecological modelling*, 200(1-2), pp.130-138.
- Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer

- Luzopone C.A., Knight R. 2007. Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* 104:11436 –11440.
- Lynch M.D., Neufeld J.D., 2015. Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* 2015;13:217-29.
- Lynch, J.M. and Bragg, E., 1985. Microorganisms and soil aggregate stability. In *Advances in soil science* (pp. 133-171). Springer, New York, NY.
- Madsen, E.L., 2011. Microorganisms and their roles in fundamental biogeochemical cycles. *Current opinion in biotechnology*, 22(3), pp.456-464.
- Manale, A, Sharpley A, DeLong C, et al., 2018. Principles and policies for soil and water conservation. *Journal of Soil and Water Conservation*. 73(4), pgs. 96A-99A
- Mangiarotti, S., Sekhar, M., Berthon, L., Javeed, Y. and Mazzega, P., 2012. Causality analysis of groundwater dynamics based on a Vector Autoregressive model in the semi-arid basin of Gundal (South India). *Journal of Applied Geophysics*, 83, pp.1-10.
- Martiny JBH, Bohannan BJ, Brown JH et al., 2006. Microbial biogeography: Putting microorganisms on the map. *Nat Rev Microbiol*;4:102-12.
- Masih, AMM., Masih, R., 1996. Energy consumption, real income and temporal causality: results from a multi-country study based on cointegration and error-correction modeling techniques. *Energy Economics* 18, 165-183.
- Matsukawa, S. and Wada, T.A.K.A.O., 1997. Vector autoregressive modeling for analyzing feedback regulation between heart rate and blood pressure. *American Journal of Physiology-Heart and Circulatory Physiology*, 273(1), pp.H478 H486.
- McClain, ME, Boyer, EW., Dent, CL., 2003. Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystem. *Ecosystems*, 6, 301-312.
- Mestre, M., Ruiz-González, C., Logares, R., Duarte, C.M., Gasol, J.M. and Sala, M.M., 2018. Sinking particles promote vertical connectivity in the ocean microbiome. *Proceedings of the National Academy of Sciences*, 115(29), pp.E6799-E6807.

- Mitchell, AC,., Dideriksen, K,., Spangler, LH, Cunningham AB, Gerlach R., 2010., et. al., 2010. Microbially enhanced carbon capture and storage by mineral-trapping and solubility-trapping. *Environ. Sci. Technol* 44:, 5270-5276 <http://doi.org/>. DOI: 10.1021/es903270w
- Montanaro, L., Poggi, A., Visai, L., Ravaioli, S., Campoccia, D., Speziale, P. and Arciola, C.R., 2011. Extracellular DNA in biofilms. *The International journal of artificial organs*, 34(9), pp.824-831.
- Mullen K (2012) Information on Earth's water. NGWA www.ngwa.org
- Musslewhite CL, McInerney MJ, Dong H et al. The factors controlling microbial distribution and activity in the shallow subsurface. *Geomicrobiol* 2003;20: 245-61.
- Naeem, S., Li, S. 1997. Biodiversity enhances ecosystem reliability. *Nature* 390, 507-509.
- Nakajima, T., 2014. Price Formation of Coffee Beans: New Evidence from the Japanese Market. *Research in World Economy*, 5(2), pp.23-30.
- National Research Council, Division on Earth and Life Studies, Commission on Geosciences, Environmental and Resources., 2000. *Research Needs in Subsurface Science*. National Academies Press.
- Nelson, D., 2002. Natural variations in the composition of groundwater. *Drinking Water Program*. Oregon Department of Human Services, Springfield, Oregon, 3. Nelson, D. (2002). Natural variations in the composition of groundwater. Presented at Groundwater Foundation Annual Meeting, November.
- Nemergut DR et al., 2011. Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* 13: 135-144
- Nemergut DR et al., 2013. Patterns and Processes of Microbial Community Assembly. *Microbiol Mol Biol Rev* 77: 342-356
- Nielsen, K.M., Johnsen, P.J., Bensasson, D. and Daffonchio, D., 2007. Release and persistence of extracellular DNA in the environment. *Environmental biosafety research*, 6 (1-2), pp.37-53.
- Nielsen, T.H., Sjøholm, O.R. and Sørensen, J., 2009. Multiple physiological states of a *Pseudomonas fluorescens* DR54 biocontrol inoculant monitored by a new flow cytometry protocol. *FEMS microbiology ecology*, 67(3), pp.479-490.

- Nimmo, J.R., 2004, Porosity and Pore Size Distribution, in Hillel, D., ed. Encyclopedia of Soils in the Environment: London, Elsevier, v. 3, p. 295-303.
- Nocker, A., Cheung, C.Y. & Camper, A.K. 2006. Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from the cells. *Journal of Microbiological Methods*, 67, 310–320.
- Nocker, A., Sossa-Fernandez, P., Burr, M.D. and Camper, A.K., 2007. Use of propidium monoazide for live/dead distinction in microbial ecology. *Appl. Environ. Microbiol.*, 73(16), pp.5111-5117.
- O'Malley MA. The nineteenth century roots of ‘Everything is everywhere’. *Nat Rev Microbiol* 2007;5:647.
- Oksanen J, Guillaume Blachet F, Kindt R, et al., 2013. *Vegan: Community Ecology Package*. R Package Version. 2.0-10. <https://cran.r-project.org>, <https://github.com/vegandevs/vegan>
- Oksanen J, Guillaume Blanchet F, Friendly M, et al., 2017. *vegan: Community Ecology Package*. R package version 2.4-3. <https://CRAN.R-project.org/package=vegan>
- Oksanen, J., Blanchet, FG., Kindt, R., et al., 2013. *Vegan: Community Ecology Package*. R-package version 2.0-10.
- Or D, Smets BF, Wraith JM et al., 2007. Physical constraints affecting bacterial habitats and activity in unsaturated porous media – a review. *Adv Water Resour*; 30:1505–27.
- Papapetrou, E., 2001. Oil price shocks, stock market, economic activity and employment in Greece. *Energy economics*, 23(5), pp.511-532.
- Pedrós-Alió, C., 2012. The rare bacterial biosphere. *Ann. Rev. Mar. Sci.* 4, 449 – 466.
- Pepper IL, Brusseau ML. Physical-chemical characteristics of soils and the subsurface. In: Pepper IL et al. (eds.). *Environmental and Pollution Science*, 2nd Edition. San Diego, CA: Elsevier Science/Academic Press, 2006, 13–23.
- Pester M, Bittner N, Deevong P, et al. A ‘rare biosphere’ microorganism contributes to sulfate reduction in a peatland. *ISME J* 2010;4:1-12.
- Pfaff, B., 2007. Package “vars”. <http://cran.r-project.org/web/packages/vars/index.html>

- Pfaff, B., 2008. Analysis of integrated and cointegrated time series with R. Springer. 10.1007/978-0-387-75967-8
- Pietramellara G. et al., 2008. Extracellular DNA in soil and sediment: fate and ecological relevance. *Biol Fertil Soils*. 45, 219–235
- Preheim, SP., Perrotta, AR., Martin-Platero, AM., et. al., 2013. Distribution-based clustering: Using ecology to refine the operational taxonomic unit. *Appl. Environ. Microbiol* 79 (21), 6593-6603.
- Price, MN., Dehal, PS., Arkin, AP., 2009. FastTree: computing large minimum evolution trees with Profiles instead of a distance matrix. *Molecular Biology and Evolution* 26 (7): 1641–1650.
- Price, MN., Dehal, PS., Arkin, AP., 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5 (3), e9490. <https://doi.org/10.1371/journal.pone.0009490>
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 35(2) pgs 7188-96.
- R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rabbi, S.M., Tighe, M.K., Flavel, R.J., Kaiser, B.N., Guppy, C.N., Zhang, X. and Young, I.M., 2018. Plant roots redesign the rhizosphere to alter the three-dimensional physical architecture and water dynamics. *new phytologist*, 219(2), pp.542-550.
- Rabbi, S.M.F., Daniel, H., Lockwood, P.V., Macdonald, C., Pereg, L., Tighe, M., Wilson, B.R. and Young, I.M., 2016. Physical soil architectural traits are functionally linked to carbon decomposition and bacterial diversity. *Scientific reports*, 6, p.33012.
- Ramírez, G.A., Jørgensen, S.L., Zhao, R. and D'Hondt, S., 2018. Minimal influence of extracellular DNA on molecular surveys of marine sedimentary communities. *Frontiers in microbiology*, 9.
- Reardon, C.L., Cummings, D.E., Petzke, L.M., Kinsall, B.L., Watson, D.B., Peyton, B.M. and Geesey, G.G., 2004. Composition and diversity of microbial communities recovered from surrogate minerals incubated in an acidic uranium contaminated aquifer. *Appl. Environ. Microbiol.*, 70(10), pp.6037-6046.

- Redford, A.J., Fierer, N. (., 2009).. Bacterial succession on the leaf surface: a novel system for studying successional dynamics. *Microb Ecol* 58:, 189–198.
- Reese, A.T. and Dunn, R.R., 2018. Drivers of microbiome biodiversity: a review of general rules, feces, and ignorance. *Mbio*, 9(4), pp.e01294-18.
- Reilly, TE., Dennehy, KF., Alley, WM., Cunningham, WL., 2008. Groundwater availability in the United States. U.S. Geological Survey Circular 1323, 70p. <http://pubs.usgs.gov/circ/1323/>
- Rehfeld, K., Marwan, N., Heitzig, J. and Kurths, J., 2011. Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes in Geophysics*, 18(3), pp.389-404.
- Ruamps, LS; Nunan, N., Chenu C., 2011. Microbial biogeography at the soil pore scale. *Soil Biology and Biochemistry*. 43(2), pg. 280-286
- Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F., 2006. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* 22(20): 2532-2538
- Rulli, M.C., Savori, A. and D’Odorico, P., 2013. Global land and water grabbing. *Proceedings of the National Academy of Sciences*, 110(3), pp.892-897.
- Shade, A., Kent, AD., Jones, SE., 2007. Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol Oceanogr* 52, 487–494.
- Shade, A., McManus, PS., Handelsman, J., 2013a. Unexpected diversity during community succession in the apple flower microbiome. *mBio* 4, e00602–e00612.
- Shade, A., Caporaso, JG., Handelsman, J., 2013b. A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J* 7 (8), 1493-506.
- Shade, A., Jones, SE., Caporaso, JG., et al., 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* 5, e01371-14.
- Shannon P, Markeil A, Ozier O, et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 11: 2498-504.
- Shoemaker WR, Locey KJ, Lennon JT., 2017. A macroecological theory of microbial biodiversity. *Nat Ecol Evo*;1:0107. DOI: 10.1038/s41559-017-0107.

- Sievers, F., Wilm, A., Dineen, D., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7, 539-539.
- Silliman SE, Berkowitz B, Simunek J et al., 2002. Fluid flow and solute migration within the capillary fringe. *Groundwater*; 40:76-84.
- Simon, K.S., J. Gilbert, P. Petitot, and R. Laurent. 2001. Spatial and temporal patterns of bacterial density and metabolic activity in a karst aquifer. *Arch. Hydrobiologie* 151:67-82.
- Smith, H.J., A.J. Zelaya, K.B. De León, et al., 2018. Impact of hydrologic boundaries on microbial planktonic and biofilm communities in shallow terrestrial subsurface environments. *FEMS Microbiol. Ecol* 94 (12), fyy191.
- Smith, H.J., Foster, R.A., McKnight, D.M., Lisle, J.T., Littmann, S., Kuypers, M.M. and Foreman, C.M., 2017. Microbial formation of labile organic carbon in Antarctic glacial environments. *Nature Geoscience*, 10(5), p.356.
- Sogin, ML., Morrison, HG., Huber, JA., et al., 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA* 32, 12115-12120.
- Solomon, DK., Moore, GK., Toran, LE., Dreier, RB., McMaster, WM., 1992. Status report: A hydrologic framework for the oak ridge reservation. DOE Environmental Sciences Division. https://www.researchgate.net/publication/236370038_
- Staley, C. and Sadowsky, M.J., 2016. Regional similarities and consistent patterns of local variation in beach sand bacterial communities throughout the northern hemisphere. *Appl. Environ. Microbiol.*, 82(9), pp.2751-2762.
- Stegen JC, Lin X, Konopka AE, Fredrickson JK., 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 6: 1653-1664.
- Stegen, JC., Lin, X., Fredrickson, JK., et al., 2013. Quantifying community assembly processes and identifying features that impose them. *ISME J* 7(11), 2069-79.
- Stegen, JC., Lin, X., Fredrickson, JK., Konopka, AE., 2015. Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers Microbiol.* 6:370.
- Sui, Q., Cao, X., Lu, S, Zhao, W, Qiu, Z., Yu, G., 2015. Occurrences, sources and fate of pharmaceuticals and personal care products in the groundwater: A review. *Emerging Contaminants* 1, 14-24.

- Sun, C.S., Wang, Y.N. and Li, X.R., 2008. A vector autoregression model of hourly wind speed and its applications in hourly wind speed forecasting. *Proceedings Chinese Society of Electrical Engineering*, 28(14), p.112.
- Thakur, M.P. and Wright, A.J., 2017. Environmental filtering, niche construction, and trait variability: The missing discussion. *Trends in ecology & evolution*, 32(12), pp.884-886.
- Tilman, D., Balzer, C., Hill, J. and Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, 108(50), pp.20260-20264.
- Toth J. A theoretical analysis of groundwater flow in small drainage basins. *Geophys Res* 1963;68:4795-812.
- Trapletti A., Hornik, K., 2018. Tseries: Time Series analysis and computational finance. R package version 0.10-46. <https://CRAN.R-project.org/package=tseries>
- United Nations World Water Assessment Programme (UNWWAP)., 2018. The United Nations World Water Development Report: Water for a Sustainable World (UNESCO, Paris), Vols 1–8.
- van der Gun, J., Aureli, A., Merla, A. 2016. Enhancing groundwater governance by making the linkage with multiple uses of the subsurface space and other subsurface resources. *Water*, 8, 222.
- Van Ree, C.C.D.F. and Van Beukering, P.J.H., 2016. Geosystem services: a concept in support of sustainable development of the subsurface. *Ecosystem services*, 20, pp.30-36.
- Veresglou SD, Halley JM, Rillig MC., 2015. Extinction risk of soil biota. *Nat Comm* 6: 1-10
- Vroblesky, DA., Chapelle, FH., 1994. Temporal and spatial changes of terminal electron-accepting processes in a petroleum hydrocarbon-contaminated aquifer and the significance for contaminant biodegradation. *Water Resources and Res.* 30,1561-1570.
- Wall, D.H., Bardgett, R.D., Covich, A.P. and Snelgrove, P.V., 2004. The need for understanding how biodiversity and ecosystem functioning affect ecosystem services in soils and sediments (pp. 1-12). Island Press: Washington, DC, USA.
- Waller, R.M. 1988. Ground Water and the Rural Homeowner, USGS. https://pubs.usgs.gov/gip/gw_ruralhomeowner/pdf/gw_ruralhomeowner.pdf

- Wang, Q., Garrity, GM., Tiedje, JM., Cole, JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73 (16), 5261-5267.
- Watson, DB., Kostka, JE., Fields, MW., Jardine, PM., 2004. The Oak Ridge Field Research Center Conceptual Model. DOE Environmental Sciences Division. <https://public.ornl.gov/orifc/FRC-conceptual-model.pdf>
- Webb, CO., Ackerly, DD., Kembel, SW., 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24, 2098-2100.
- Weiss S, Van Treuren W, Lozupone C, et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10:7, pgs. 1669-1681.
- Wertz, S., Degrange, V., Prosser, JI., et al., 2007. Decline of soil microbial diversity does not influence the resistance and resilience of key soil microbial functional groups following a model disturbance. *Environ Microbiol* 9, 2211–2219.
- Whitman T., Neurath R., Perera A., Chu-Jacoby I., Ning D., Zhou J., Nico P., Pett-Ridge J., and Firestone M. Microbial community assembly differs across minerals in a rhizosphere microcosm. *Environmental Microbiology* 20, 4444-4460 (2018) doi:doi:10.1111/1462-2920.14366.
- Whitman, W.B., Coleman, D.C. and Wiebe, W.J., 1998. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, 95(12), pp.6578-6583.
- Wolkovich, E.M., Cook, B.I., McLauchlan, K.K. and Davies, T.J., 2014. Temporal ecology in the Anthropocene. *Ecology letters*, 17(11), pp.1365-1379.
- Wu, L., Liu, X., Schadt, C.W., and Zhou, J., 2006. Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Appl Environ Microbiol* 72, 4931–4941.
- Yabusaki, SB., Wilkins, MJ., Fang, Y, et al., 2017. Water table dynamics and biogeochemical cycling in a shallow, variably-saturated floodplain. *Environ. Sci. Technol.* 51, 3307–3317.
- Yan Q., Li J., Yu Y., Wang J., He Z., Van Nostrand J. D., Kemper M. L., Wu L., Wang Y., Liao L., Li X., Wu S., Ni J., Wang C., and Zhou J., 2016. Environmental filtering decreases with fish development for the assembly of gut microbiota. *Environmental Microbiology* 18, 4739-4754. DOI:10.1111/1462-2920.13365.

- Yan Q., Stegen J. C., Yu Y., Deng Y., Li X., Wu S., Dai L., Zhang X., Li J., Wang C., Ni J., Li X., Hu H., Xiao F., Feng W., Ning D., He Z., Van Nostrand J. D., Wu L., and Zhou J., 2017. Nearly a decade-long repeatable seasonal diversity patterns of bacterioplankton communities in the eutrophic Lake Donghu (Wuhan, China). *Molecular ecology* 26, 3839-3850. DOI:10.1111/mec.14151.
- Zelaya AJ et al (submitted). High spatiotemporal variability of bacterial diversity over short time scales with unique geohydrochemistry within a shallow, pristine aquifer. *Water Research*.
- Zhang C, Palumbo AV, Phelps TJ et al., 1998. Grain size and depth constraints on microbial variability in coastal plain subsurface sediments. *Geomicrobiol J*; 15:171–85.
- Zhou J, Ning D. 2017. Stochastic community assembly: does it matter in microbial ecology? *Microbiol Mol Biol Rev* 81:e00002-17. <https://doi.org/10.1128/MMBR.00002-17>.
- Zhou Y, Kellermann C, Griebler C., 2012. Spatio-temporal patterns of microbial communities in a hydrologically dynamic pristine aquifer. *FEMS Microbiol Ecol* 81, 230-242.
- Zinger, L., Gobet, A., Pommiers, T., 2012. Two decades of describing the unseen majority of aquatic microbial diversity. 21, 1878-1896.